5-8-2020

Speciation Genomics of the Chrysoperla carnea Complex

Katherine Taylor
University of Connecticut - Storrs, katherine.l.taylor@uconn.edu

Speciation genomics of the *Chrysoperla carnea* complex

Katherine L Taylor, PhD

University of Connecticut, 2020

Rapid species radiations can provide insight into the process of speciation and diversification. *Chrysoperla carnea,* the common green lacewing, was once thought to be a single insect species with a near worldwide distribution. Species-specific vibrational mating songs revealed more than twenty morphologically cryptic species in the *Chrysoperla carnea*-group. Rapid diversification in this clade seems to have been driven, at least in part, by their precise substrate-borne vibrational duets performed prior to copulation. In this dissertation, I examine speciation in the *Chrysoperla carnea*-group by reconstructing the evolutionary history of the clade and identifying the genomic basis of a mating song trait critical to the maintenance of species boundaries in this group.

In Chapter 1, I assemble and annotate a *de novo* reference genome of *Chrysoperla carnea* sensu stricto, the first available for a neuropteran insect. In Chapter 2, I infer the evolutionary history of the *carnea*-group using reduced representation genome sequencing. Resolved species relationships indicate repeated phenotypic radiations, with multiple parallel evolution of ecomorphs on different continents. These analyses also reveal a history of hybridization in the clade. In Chapter 3, I identify a single large genomic region associated with the mating song feature 'volley period,' using QTL mapping. This genomic region contains several candidate genes for lacewing song phenotype, including the gene *doublesex*, which is critical to song production in *Drosophila*. Additionally, I demonstrate strong genetic linkage between volley period phenotype and preference for volley period. Overall, this dissertation advances our understanding of *carnea*-group lacewing diversification and speciation.

Speciation genomics of the *Chrysoperla carnea* complex

Katherine L Taylor

B.S., Mount Holyoke College, 2012

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy
at the
University of Connecticut

2020

APPROVAL PAGE

Doctor of Philosophy Dissertation

Presented by

Katherine L Taylor, B.S.

Major advisor _____

Charles Henry

Associate advisor_____

Elizabeth Jockusch

Associate advisor_____

Chris Simon

Associate advisor_____

Jill Wegrzyn

Associate advisor_____

David Wagner

University of Connecticut

2020

Acknowledgments

Without the support and mentorship of my major advisor, Dr. Charles Henry, and committee members Drs. Elizabeth Jockusch, Chris Simon, Jill Wegrzyn, and David Wagner, this work would not have been possible. Support from the ecology and evolutionary biology (EEB) faculty, staff, and students, and from the international lacewing community, was central to the success of the project. I would especially like to thank Drs. Marta M. Wells, Paul Lewis, Peter Duelli, Elizabeth Wade, and Atsushi Mochizuki for their critical contributions to these studies. Finally, a huge thank you to my family, my friends, and Bonnie.

Table of Contents

Approval page.....iii

Acknowledgements.....iv

Chapter 1

Draft genome assembly for the common green lacewing *Chrysoperla carnea*.....1

Figures.....10

References 17

Chapter 2

Evolutionary history of the *Chrysoperla carnea* species group.....23

Tables and Figures.....37

References 40

Chapter 3

Candidate genomic region for species-specific mating song in lacewings.....49

Figures.....59

References 63

Draft genome assembly for the common green lacewing *Chrysoperla carnea*

Katherine L Taylor, Elizabeth Wade, Charles Henry, Jill Wegrzyn

Context:

Neuroptera is a holometabolous insect order comprising the lacewings, antlions, owlflies, and some 14 other families. This diverse order, found nearly worldwide, encompasses 5,813 described species (Oswald & Machado, 2018).Yet, the only nuclear genomic resources available for neuropteran insects are transcriptomes (Li et al., 2013). The present project constructed and annotated the first nuclear genome assembly to assist investigations related to the biology of the Neuroptera broadly and more specifically on the common green lacewing *Chrysoperla carnea* (Stephens) (NCBI:txid189513) (Fig. 1). The common green lacewing is among the most extensively studied insects in the order Neuroptera, due to its use as a non-model evolutionary study system and its importance as a natural enemy of arthropod agricultural pests.

Larvae of *Chrysoperla* are important generalist predators used in the biological control of agricultural pests including aphids, whiteflies, and other soft-bodied arthropods (Senior et al., 2001). Their oligophagous nature, amenability to mass rearing and release, and resistance to multiple classes of insecticides make them well suited to use in integrated pest management plans (Bielza, 2016; Pappas et al., 2011). There are active research programs investigating mechanisms underlying pesticide resistance in these insects, motivated by the more than 160 reports of *Chrysoperla carnea* populations developing resistance to 22 different active compounds found in insecticides (Mansoor & Shad, 2019; Mota-Sanchez & Wise, 2019; Venkatesan et al., 2017). Understanding the genetic basis of pesticide resistance is one way the genomic resource presented here might be used to improve the biological control of insect pests. The genome might also be useful for elucidating the genetic basis of geographically varying life

history traits relevant to their role as predators such as dietary preference and thermal tolerance (Henry & Wells, 2007).

These lacewings have also been used to address varied questions in evolution, ecology, and behavior. *Chrysoperla carnea* is a member of a cryptic lacewing species complex, which has recently and rapidly radiated apparently due to isolation by pre-mating substrate-borne vibrational duets (Henry et al., 2013). Neuroptera is one of few insect groups with described widespread use of this communication form (Virant-Doberlet & Cokl, 2004). Seemingly because of signal-mediated isolation the species closely related to *Chrysoperla carnea,* despite being interfertile, have evolved differences in life cycle timing and ecological niche (Tauber & Tauber, 1977; Thierry et al., 2011). Their amenability to mass rearing, ability to produce fertile hybrids, and 1-month generation time make them a tractable emerging model system. With appropriate genomic resources for *Chrysoperla carnea* researchers could address the genomic architecture of speciation (Noh & Henry, 2015), the evolutionary history of Chrysopidae and *Chrysoperla*, how selective and non-selective forces shape the genome (Henry & Wells, 2004), and the genetics of life-cycle switches (Duelli et al., 2014; Tauber & Tauber, 1977), ecological specialization (Thierry et al., 2011), and mating song diversity (Henry et al., 2013).

Methods and results:

Sample preparation and sequencing:

The insects were sourced from an inbred lab-reared colony of *Chrysoperla carnea*, established with wild-caught individuals from Zürich, Switzerland. All sequence library preparation and sequencing was completed by the Interdisciplinary Center for Biotechnology Research at the University of Florida. DNA of a single individual, extracted with a DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany), was used to construct the paired end (PE) Illumina DNA library

with an average insert size of 413 base pairs (bp). The library was sequenced on one lane of an Illumina HiSeq 4000 (Illumina Inc., San Diego, CA, USA) which resulted in over 330 million 150 bp PE reads. DNA from multiple siblings was pooled after whole body extraction to yield sufficient quantity for SMRT bell library preparation. The long read library was SMRT sequenced on four cells of a Pacific Biosciences Sequel System (Pacific Biosciences, Menlo Park, CA, USA). Long read sequencing efforts produced 307,542 polymerase reads with a mean length of 9.5 Kb.

Nuclear genome assembly:

Adaptor contamination and low quality regions were trimmed from the Illumina short reads using ILLUMINACLIP:2:30:10:4, SLIDINGWINDOW:4:20, and MINLEN:45 functions of Trimmomatic 0.36 (Bolger et al., 2014). Contaminant reads were filtered based upon similarity to bacterial and fungal genomes on NCBI RefSeq with Kraken 1.0 (Wood & Salzberg, 2014). After quality control, 214 million 150 bp PE reads remained for assembly, yielding theoretical 100x genome coverage.

Genome characteristics were estimated based on k-mer counts produced by the BBtools kmercountexact script (Bushnell, n.d.) for k-mer sizes 21, 23, 25, 27, and 29 (Fig. 2). The average genome size estimated from the trimmed reads was 645 MB while the average heterozygosity was 1.7%.

Short reads were assembled using four software packages. The MaSuRCA 3.2.4 (Zimin et al., 2013) assembler, a combined deBruijn graph and Overlap-Layout-Consensus hybrid approach, was run on the untrimmed short reads with default parameters except cgwErrorRate=0.10 and resource allocation parameters. ABySS 2.0.2 (Simpson et al., 2009) and SOAPdenovo 2.04 (Luo et al., 2012), two deBruijn graph short read assemblers for moderately

sized eukaryotic genomes, were run with default parameters on the trimmed short read set at increasing k-mer sizes until assembly statistics (length and N50) plateaued; the k-mer sizes used ranged from 37 –121. SPAdes 3.11.1 (Bankevich et al., 2012), a deBruijn graph short read assemblers designed for bacterial and other small genomes was run with default parameters on the trimmed short read set with default settings and k-mer sizes 37, 53, 71, 93, 101, 107, 113, 121, and 127. All assemblies were assessed with Quast (4.2) (Mikheenko et al., 2016). The longest and most contiguous assembly from each assembler was selected for scaffolding and heterozygosity reduction (**Table 1**).

The low coverage (<5x) PacBio long reads were prepared for scaffolding with circular consensus sequencing (CCS) for error rate reduction; CCS reads were generated from at least five sequencing passes and with predicted base calling accuracy of at least 0.9 in the PacBio SMRT portal version 7. A total of 122 CCS reads with an average length of 4.9Kb resulted. The reads were compared to the RefSeq database via BLAST 2.7.1 (Camacho et al., 2009) to identify contaminant sequences. The bacterial, fungal, and viral reference assemblies with the most BLAST alignments (> 50 bp and E-values < 0.001) were downloaded. The metaquast.py script from Quast 5.0.2 (Mikheenko et al., 2015) detected alignments greater than 150 bp between the CCS reads and the potential contaminant genomes. A total of 105 reads, with an average length of 4.1Kb and estimated coverage of < 1x, remained for scaffolding.

The best assembly from each of the short read assemblers was heterozygosity reduced and scaffolded with the CCS long reads using the Redundans pipeline (Pryszcz & Gabaldón, 2016). Homologous contigs from separately assembled heterozygous regions were identified from read depth and alignment similarity, and the smaller of the redundant contigs were removed. Iterative mapping of the CCS and short reads to the reduced assembly provided the

scaffolding input for SSPACE3 (Boetzer et al., 2010). Following scaffolding, gaps were filled with GapCloser (Luo et al., 2012).

The scaffolded assemblies were evaluated with Quast 4.2 (Mikheenko et al., 2016). The Redundans pipeline decreased the length and increased the N50 for all assemblies (**Table 1**). The assembly produced by MaSuRCA and heterozygosity-reduced and scaffolded by Redundans was the most contiguous of the assemblies and closest to the estimated genome size; it therefore was selected for subsequent annotation. Before annotation, scaffolds with complete or partial origin from contaminants or the mitochondrial genome were removed following the procedures described above for the CCS reads. The final assembly was 616 Mb with an N50 of 26.1Kb represented by 40,501 contigs. Conserved single-copy orthologous genes in the BUSCO insecta_odb9 database were identified with BUSCO 3.0.2 using the tribolium2012 gene finding parameters for Augustus (Waterhouse et al., 2017). Of the 1658 genes in the database, 95.2% were found compete in the assembly (**Table 2**).

Nuclear genome annotation:

A consensus library of interspersed repetitive content was constructed with RepeatModeler 1.0.8 (Smit & Hubley, 2008) Alignment of this library with RepeatMasker 4.0.6 (Smit et al., 2004) softmasked 40.5% of the reference genome. Two publicly available RNA-Seq data sets from very closely related species were aligned using default parameters to the masked final assembly using HiSat 2.1.0 (Kim et al., 2015). The first, BioProject PRJNA324742, contained 100 million 150 bp pe reads from whole body extracted RNA from larvae of *Chrysoperla zastrowi sillemi* (Esben-Petersen) (NCBI:txid482137) sequenced on an Illumina HiSeq 2500, of which 67% of the reads aligned to the reference genome. The second, 64 million 90 bp pe reads from BioProject PRJNA267467, from *Chrysoperla nipponensis* (Okamoto)

(NCBI:txid413239) adult antenna sequenced on Illumina HiSeq 2000, resulted in a 59% read alignment rate.

BRAKER 2.0.5 (Hoff et al., 2015) was used to predict gene space and features using RNA-Seq evidence with GeneMark-ET (Lomsadze et al., 2014) and Augustus (Stanke et al., 2006). BRAKER identified 30,670 gene models. These gene models were annotated with the EnTAP pipeline (Hart et al., 2018), which mines for functional information from multiple databases and lines of evidence. Homology between translated proteins from the gene models and the annotated proteins in the RefSeq and UniProt databases was determined with a similarity search by DIAMOND (Buchfink et al., 2015). Alignments with E-values < .0001 that covered 60% or more protein length were considered homologous and used for functional annotation. Similarity searching produced potential annotations for 16,202 genes and flagged 125 gene models as potential bacterial or fungal contaminants. Gene families, gene ontology terms, and protein domains were identified for the gene model protein set by EggNOG-mapper (Huerta-Cepas et al., 2017) hits to the EggNOG database and InterProScan 74.0 (Jones et al., 2014) hits to the Pfam and SMARTdatabases. This process resulted in 24,363 genes with identified family assignments, gene ontology terms, or protein domains. Gene models that were both (i) without at least one identified protein domain, gene ontology term, or family assignment and (ii) beginning or ending with an intron, lacking start or stop codon, or with more than one in-frame stop codon or (iii) EnTAP-flagged contaminant genes were removed from the gene set. Of the 30,670 gene models produced by BRAKER, 28,842 genes models were retained in the final gene set. In the final set of 28,842 genes, 85.3% are complete. Gene statistics were calculated by gFACs (Caballero & Wegrzyn, 2019) (**Table 3**).

Mitochondrial genome assembly and annotation

The mitochondrial genome was assembled by iterative mapping of the trimmed reads by MITObim (Hahn et al., 2013) using the *Chrysoperla nipponensis* mitochondrial genome accession KP027407 (Haruyama et al., 2011) as a reference. The assembly was circularized by identifying overlap between the two ends of the contig and correct circularity was confirmed by mapping reads back across the joined ends using AWA (Jacob Machado et al., 2018). This resulted in a 16 Kb circular genome with 21.2% GC content (**Table S1**). This assembly covers the full length with 98.5% similarity to the closely related *Chrysoperla nipponensis* mitochondrial genome accession KP027407 (Haruyama et al. 2011). Notably, this assembly has a gene order shift from the standard insect mitochondrial genome (trnW-trnC-trnY → trnC-trnW-trnY) which has been found in the mitochondrial genomes of other neuropteran insects (Beckenbach & Stewart, 2008).

Comparative genomic assessment:

Orthogroups were inferred for the *Chrysoperla carnea* gene set, a whole body transcriptome from a neuropteran insect *Chrysopa pallens* (Rambur) *(*NCBI:txid 417485 BioProject PRJNA186574), and full-length representative genome assemblies of seven coleopterans, Neuroptera's closest relative with genomic resources. The beetle gene sets were, *Tribolium castaneum* (ID: 216 NCBI:txid7070), *Nicrophorus vespilloides* (ID: 40824 NCBI:txid110193), *Onthophagus taurus* (ID: 12827 NCBI:txid166361), *Dendroctonus ponderosae* (ID: 11242 NCBI:txid77166), *Anoplophora glabripennis* (ID: 14033 NCBI:txid217634), *Agrilus planipennis* (ID: 12835 NCBI:txid224129), and *Diabrotica virgifera* (ID: 17855 NCBI:txid 50389). Orthogroups were identified with Orthofinder 2.1.2 (Emms & Kelly 2018). In neuropteran resources, relative to the coleopteran genomes, 199 orthogroups

were expanded, 50 were contracted, 1860 were missing, and 1,267 were novel (Figure 3). Using the longest *C. carnea* gene in each group, gene ontology terms were associated with orthogroups.

To explore the genomic basis of substrate-borne vibrations courtship in Neuroptera, gene ontology terms within mating behavior were identified in orthogroups that were expanded, contracted, or novel in both of the neuropteran resources. Courtship behavior terms were associated with two of the expanded orthogroups. One of these expanded orthogroup was associated with GO:0016545-male courtship behavior, veined wing vibration. For both of these expanded orthogroups associated with mating behavior six of seven beetles had a single gene in the orthogroup while each of the neuropteran resources had two, suggesting either duplication in Neuroptera or gene loss in Coleoptera.

Conclusions

We provide a high quality draft genome for *Chrysoperla carnea.* The assembly is 95.5% of the estimated length and contains 95% of the BUSCO genes in the insect database. For a primarily short read derived assembly, it is quite contiguous, although additional high coverage long reads could improve it. 83.1% of the gene models are complete and 86.4% of genes are annotated suggesting a good quality draft annotation. This reference can facilitate the use of reduced-representation or lower coverage sequencing approaches in future studies, potentially addressing questions on the biology of the Neuroptera, the improved application of *Chrysoperla* lacewings to biological control, or the impact of evolutionary processes such as speciation and natural selection in this emerging model organism.

We applied this resource to explore orthogroups associated with courtship and mating behavior, and identified two promising targets seemingly expanded in Neuroptera relative to Coleoptera. This expansion may be related to the critical importance of substrate-borne

vibrational signals for successful courtship and copulation in many neuropterans. More robust comparative genomic analysis will be possible as more genomes are assembled and annotated for the Neuroptera and their relatives.

Tables and figures

Table 1. Statistics calculated with Quast on the contigs of 500 bp or larger for best assembly from each assembler before and after redundans scaffolding and heterozygosity reduction, and the final assembly after contaminant and mitochondrial scaffold removal.

	k-mer	N50	Length	# Contigs
ABySS	97	6,536	696,784,456	201,286
ABySS Redundans	-	8,868	562,798,940	122,075
SOAPdenovo2	113	7,319	724,434,118	198,067
SOAPdenovo2 Redundans	-	10,204	563,951,406	122,133
SPAdes	37 – 127	7,749	710,568,935	197,205
SPAdes Redundans	-	11,137	563,475,872	117,043
MaSuRCA	105	21,655	749,801,350	74,241
MaSuRCA Redundans	-	25,813	624,143,121	42,003
Final Assembly	-	26,110	616,666,273	40,488

Table 2. Presence of the genes from the BUSCO insect database in the final assembly.

	Count	Percent
Complete	1578	95.2%
Complete single-copy	1403	84.6%
Complete duplicated	175	10.6%
Fragmented	31	1.9%
Missing	49	2.9%

Table 3. Summary statistics on the final gene set after filtering calculated by gFACs.

Total genes	28,842
Monoexonic genes	7,184
Multiexonic genes	21,658
Complete genes	24,590
5' incomplete genes	2,348
3' incomplete genes	1,735
5' and 3' incomplete genes	169
Average gene size (bp)	4,268.6
Average exon size (bp)	303.7
Average CDS size (bp)	1242.9
canonical splice sites (%)	99.8
GC content (%)	34.4

Table S1. Mitochondrial gene features annotated by Mitos.

Feature	Start	Stop	Strand	Length
tRNA I	152	216	+	65
tRNA Q	291	359	-	69
tRNA M	362	429	+	68
NAD2	430	1443	+	1014
tRNA C	1442	1504	-	63
tRNA W	1505	1572	+	68
tRNA Y	1575	1642	-	68
COX1	1653	3188	+	1536
tRNA L2	3184	3248	+	65
COX2	3274	3960	+	687
tRNA K	3941	4011	+	71
tRNA D	4012	4076	+	65
ATP8	4077	4235	+	159
ATP6	4229	4906	+	678
COX3	4906	5694	+	789
tRNA G	5694	5758	+	65
NAD3	5759	6112	+	354
tRNA A	6111	6176	+	66
tRNA R	6178	6241	+	64
tRNA N	6241	6307	+	67
tRNA S1	6308	6374	+	67
tRNA E	6378	6445	+	68
tRNA F	6444	6509	-	66
NAD5	6490	8214	-	1725
tRNA H	8236	8299	-	64
NAD4	8299	9636	-	1338
NAD4L	9630	9926	-	297
tRNA T	9929	9992	+	64
tRNA P	9993	10058	-	66
NAD6	10072	10569	+	498
COB	10569	11705	+	1137
tRNA S2	11704	11771	+	68
NAD1	11798	12727	-	930
tRNA L1	12750	12813	-	64
rRNA L	12791	14101	-	1311
tRNA V	14121	14191	-	71
rRNA S	14190	14969	-	780

Figure 1. *Chrysoperla carnea* adult. Photograph by Dr. Peter Duelli.

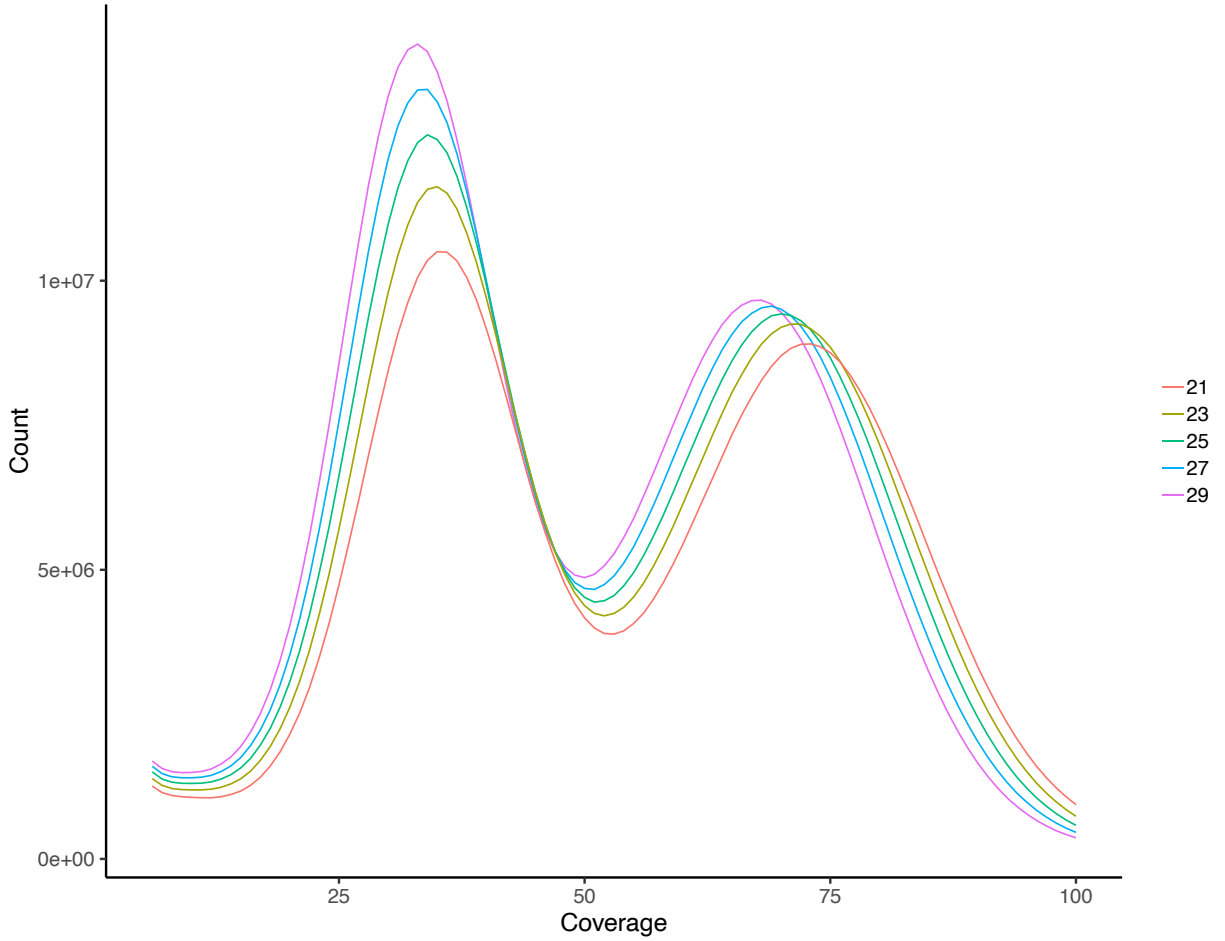

Figure 2. Illumina filtered k-mer count density. Calculated by bbmap for k-mer sizes 21, 23, 25, 27, 29. The double peaks are characteristic of genomes with high heterozygosity.

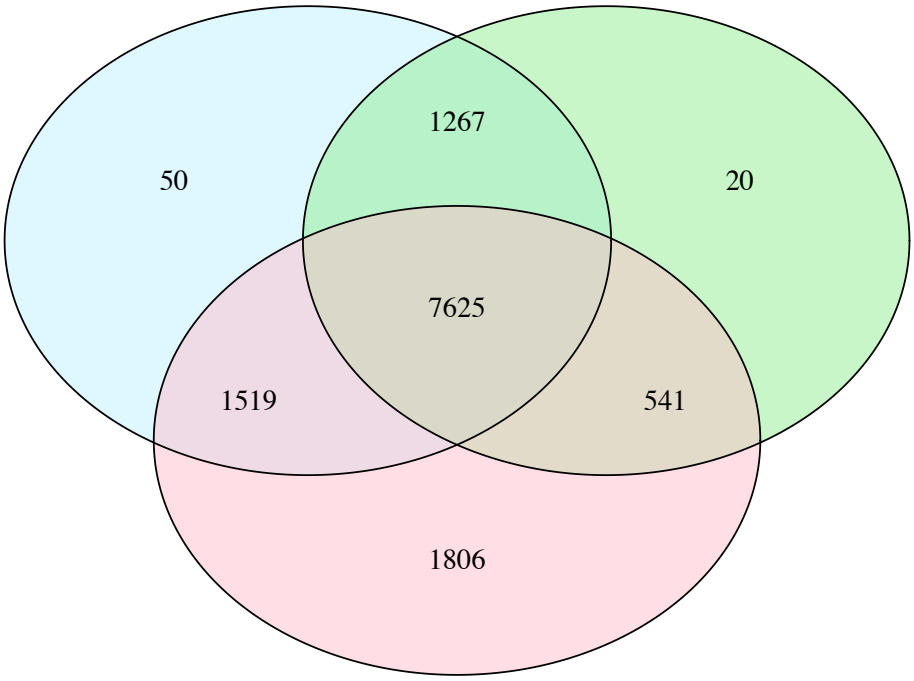

Figure 3. Orthogroup overlap between *C. carnea* genome (blue), *C. pallens* transcriptome (green), and coleopteran reference genomes (pink).

References

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., & Prjibelski, A. D. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*(5), 455–477.

Beckenbach, A. T., & Stewart, J. B. (2008). Insect mitochondrial genomics 3: The complete mitochondrial genome sequences of representatives from two neuropteroid orders: A dobsonfly (order Megaloptera) and a giant lacewing and an owlfly (order Neuroptera). *Genome*, *52*(1), 31–38.

Bielza, P. (2016). Insecticide resistance in natural enemies. In *Advances in insect control and resistance management* (Vol. 1, pp. 313–329). Springer.

Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D., & Pirovano, W. (2010). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics*, *27*(4), 578–579.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), 59.

Bushnell, B. (n.d.). *BBMap*. sourceforge.net/projects/bbmap/

Caballero, M., & Wegrzyn, J. (2019). gFACs: Gene filtering, analysis, and conversion to unify genome annotations across alignment and gene prediction frameworks. *Genomics, Proteomics & Bioinformatics*, *17*(3), 305–310.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST : architecture and applications. *BMC Bioinformatics*, *10*(1), 421.

Duelli, P., Johnson, J. B., Waldburger, M., & Henry, C. S. (2014). A new look at adaptive body coloration and color change in “common green lacewings” of the genus *Chrysoperla* (Neuroptera: Chrysopidae). *Annals of the Entomological Society of America*, *107*(2), 382–388.

Hahn, C., Bachmann, L., & Chevreux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—A baiting and iterative mapping approach. *Nucleic Acids Research*, *41*(13), e129–e129.

Hart, A. J., Ginzburg, S., Xu, M. S., Fisher, C. R., Rahmatpour, N., Mitton, J. B., Paul, R., & Wegrzyn, J. L. (2018). EnTAP: Bringing faster and smarter functional annotation to non-model eukaryotic transcriptomes. *BioRxiv*, *Journal Article*, 307868.

Haruyama, N., Mochizuki, A., Sato, Y., Naka, H., & Nomura, M. (2011). Complete mitochondrial genomes of two green lacewings, *Chrysoperla nipponensis* (Okamoto, 1914) and *Apochrysa matsumurae* Okamoto, 1912 (Neuroptera: Chrysopidae). *Molecular Biology Reports*, *38*(5), 3367–3373.

Henry, C. S., Brooks, S. J., Duelli, P., Johnson, J. B., Wells, M. M., & Mochizuki, A. (2013). Obligatory duetting behaviour in the Chrysoperla carnea-group of cryptic species (N europtera: C hrysopidae): Its role in shaping evolutionary history. *Biological Reviews*, *88*(4), 787–808.

Henry, C. S., & Wells, M. L. M. (2004). Adaptation or random change? The evolutionary response of songs to substrate properties in lacewings (Neuroptera: Chrysopidae: Chrysoperla). *Animal Behaviour*, *68*(4), 879–895.

Henry, C. S., & Wells, M. M. (2007). Can what we don't know about lacewing systematics hurt us? A cautionary tale about mass rearing and release of *"Chrysoperla carnea"* (Neuroptera: Chrysopidae). *American Entomologist*, *53*(1), 42–47.

Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M., & Stanke, M. (2015). BRAKER1: Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics*, *32*(5), 767–769.

Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., & Bork, P. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Molecular Biology and Evolution*, *34*(8), 2115–2122.

Jacob Machado, D., Janies, D., Brouwer, C., & Grant, T. (2018). A new strategy to infer circularity applied to four new complete frog mitogenomes. *Ecology and Evolution*, *8*(8), 4011–4018.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), 357–360. https://doi.org/10.1038/nmeth.3317

Li, Z. Q., Zhang, S., Ma, Y., Luo, J. Y., Wang, C.-Y., Lv, L. M., Dong, S. L., & Cui, J. J. (2013). First transcriptome and digital gene expression analysis in Neuroptera with an emphasis on chemoreception genes in *Chrysopa pallens* (Rambur). *PLoS One*, *8*(6), e67151.

Lomsadze, A., Burns, P. D., & Borodovsky, M. (2014). Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Research*, *42*(15), e119–e119.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., & Liu, Y. (2012). SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *Gigascience*, *1*(1), 18.

Mansoor, M. M., & Shad, S. A. (2019). Resistance of green lacewing, *Chrysoperla carnea* (Stephens), to buprofezin: Cross resistance patterns, preliminary mechanism and realized heritability. *Biological Control*, *129*(Journal Article), 123–127.

Mikheenko, A., Saveliev, V., & Gurevich, A. (2015). MetaQUAST: evaluation of metagenome assemblies. *Bioinformatics*, *32*(7), 1088–1090.

Mikheenko, A., Valin, G., Prjibelski, A., Saveliev, V., & Gurevich, A. (2016). Icarus: Visualizer for de novo assembly evaluation. *Bioinformatics*, *32*(21), 3321–3323.

Mota-Sanchez, D., & Wise, J. C. (2019). *The arthropod pesticide resistance database*. www.pesticideresistance.org/search.php

Noh, S., & Henry, C. (2015). Speciation is not necessarily easier in species with sexually monomorphic mating signals. *Journal of Evolutionary Biology*, *28*(11), 1925–1939.

Oswald, J. D., & Machado, R. J. (2018). Biodiversity of the Neuropterida (Insecta: Neuroptera, Megaloptera, and Raphidioptera). *Insect Biodiversity: Science and Society*, *2*(Journal Article), 627–672.

Pappas, M., Broufas, G., & Koveos, D. (2011). Chrysopid predators and their role in biological control. *Journal of Entomology*, *8*(3), 301–326.

Pryszcz, L. P., & Gabaldón, T. (2016). Redundans: An assembly pipeline for highly heterozygous genomes. *Nucleic Acids Research*, *44*(12), e113–e113.

Senior, L., McEwen, P., McEwen, P., New, T., & Whittington, A. (2001). The use of lacewings in biological control. *Lacewings in the Crop Environment*, *Journal Article*, 296–302.

Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J., & Birol, I. (2009). ABySS: a parallel assembler for short read sequence data. *Genome Research*, *19*(6), 1117–1123. https://doi.org/10.1101/gr.089532.108

Smit, A. F., & Hubley, R. (2008). *RepeatModeler*. http://www.repeatmasker.org

Smit, A. F., Hubley, R., & Green, P. (2004). *RepeatMasker*. http://www.repeatmasker.org

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. (2006). AUGUSTUS: Ab initio prediction of alternative transcripts. *Nucleic Acids Research*, *34*(Web Server issue), W435-439. https://doi.org/10.1093/nar/gkl200

Tauber, C. A., & Tauber, M. J. (1977). A genetic model for sympatric speciation through habitat diversification and seasonal isolation. *Nature*, *268*(5622), 702.

Thierry, D., Canard, M., Deutsch, B., Ventura, M. A., Lourenco, P., & Lode, T. (2011). Ecological character displacement in competing common green lacewings in Europe: A route to speciation? *Biological Journal of the Linnean Society*, *102*(2), 292–300.

Venkatesan, T., Mahiba, S. H., Jalali, S. K., Ramya, S. L., & Pratibha, M. (2017). Detection of insecticide resistance and mechanisms of resistance in field populations of *Chrysoperla zastrowi sillemi* (Neuroptera: Chrysopidae) collected from different geographical locations in India. *Journal of Biological Control*, *31*(3), 61–68.

Virant-Doberlet, M., & Cokl, A. (2004). Vibrational communication in insects. *Neotropical Entomology*, *33*(2), 121–134.

Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V., & Zdobnov, E. M. (2017). BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution*, *35*(3), 543–548.

Wood, D. E., & Salzberg, S. L. (2014). Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*, *15*(3), R46.

Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S. L., & Yorke, J. A. (2013). The MaSuRCA genome assembler. *Bioinformatics*, *29*(21), 2669–2677.

Evolutionary history of the *Chrysoperla carnea* species group

Katherine L Taylor, Elizabeth Wade, Charles S Henry

Introduction:

Rapid radiations provide insight into the origins and maintenance of biodiversity (Grant & Grant, 2011; Seehausen, 2006). When bursts of lineage splitting are recent, we can capture and interpret much of the diversification process still preserved in the genome. Model evolutionary organisms have shown us that the diversification process can proceed in many ways (Rundell & Price, 2009; Schluter, 2001).

Lineages may exploit ecological opportunity and produce adaptive forms. Well known examples of adaptive radiations include Darwin's finches, the *Anolis* lizards, and African cichlid fishes (Grant & Grant, 2011; Losos, 2011; Seehausen, 2006). Diversification with minimal ecological or morphological differentiation is also possible. Non-adaptive radiations have been described in *Plethodon* salamanders and in some land snails (Cook, 2008; Kozak et al., 2006). The evolutionary trajectories for radiating species may either be unique, or repeated as seen in *Anolis* lizards and Hawaiian stick spiders inhabiting different islands (Gillespie et al., 2018). Repeated radiations seem to be rare, most often seen among closely related taxa, poor dispersers, and inhabitants of islands or lakes (Losos, 2010).

While traditional models of speciation assume that gene flow ceases early in the process, diversification may proceed with or without introgression (Harrison & Larson, 2014). Hybridization and introgression are not rare, especially in recently diverged organisms (Mallet, 2005). Scientists studying disparate organisms have recently begun to appreciate the extent to which species diversify and persist even in the presence of gene flow, for example in the *Heliconius* butterfly and cichlid fish radiations (Malinsky et al., 2018; Martin et al., 2013).

Potentially, introgression could even be a driver of diversification (Seehausen, 2004), or a source of adaptive alleles (Pardo-Diaz et al., 2012).

The Chrysoperla carnea-group

In this work we will, for the first time, use genome-wide genetic markers to describe the diversification process in the recent and rapid radiation of the *Chrysoperla carnea* species group. Vibrational courtship songs (Henry et al., 2013), cryptic morphology (Brooks, 1994; Thierry et al., 2010), and ecological niche differences (Duelli et al., 2014) revealed a group of more than 20 morphologically cryptic sibling species in the *Chrysoperla carnea*-group (Neuroptera: Chrysopidae). Members of this cryptic species complex can be found nearly worldwide, with broad distributions across the Holarctic realm, the Indian subcontinent, and Africa (Henry et al., 2013). Their premating duets, produced by whole-body tremulation and transmitted through plant substrates, are thought to be strong barriers to introgression and the most reliable diagnostic criterion for species identification (Henry et al., 2013; Noh & Henry, 2010). Different ecological niches or allochrony further separate some of the sympatric species in this group (Tauber & Tauber, 1981; Thierry et al., 2011). Low mitochondrial sequence divergence (Price et al., 2015) suggests that the more than twenty distinct lineages in this group have diversified in the last 250,000 years (Papadopoulou et al., 2010).

Acoustic, morphological, and ecological evidence for species boundaries

The mating signals produced by lacewings in the *Chrysoperla carnea*-group are composed of multiple features that vary between species, yet are relatively invariant within each species. Within-species temporal and frequency features have been shown to vary by 12–20% and 2–12%, respectively (Henry et al., 2013). Many of these song characteristics are genetically determined, controlled by few loci of large effect (Henry et al., 2002). Lacewings will respond to

interspecific signals at low rates, yet occasional responses seem not to be sufficient to sustain the courtship duet through to copulation (Noh & Henry, 2010). No field-caught hybrids have been reported, though many of the sympatric species are syntopic and synchronic and thus opportunities for hybridization are apparent.

These song-delineated groups are morphologically cryptic and difficult to tell apart by morphology alone, even for taxonomic experts. Yet, detailed study has revealed morphological features or suites of features that have been used to distinguish many of the song-delineated groups. Morphological features, including the dilation of the base of the pretarsal claws, facial markings, and ground coloration of the body, sometimes differ among species (Brooks, 1994; Thierry et al., 2010). However, the utility of these characteristics for species determination is limited by intraspecific variation and the qualitative nature of the described traits.

Ecological associations can also discriminate among these distinct lineages. While some species dwell in and around evergreen conifers, others are associated with deciduous trees or herbaceous plants (Duelli et al., 2014; Macleod, 1967; Tauber & Tauber, 1977; Thierry et al., 2011). Species also seem to vary in life cycle timing and the cues that initiate and terminate diapause, but these traits have not been clearly linked to the song-delineated lineages (Tauber et al., 1977; M. J. Tauber & Tauber, 1973).

An unknown evolutionary history

Despite morphological, ecological, and acoustic evidence of isolated lineages in the *carnea*-group, all prior attempts to reconstruct the phylogenetic history of the group fail to provide support for expected species boundaries. Analyses with mitochondrial sequence data that include multiple individuals of each species suggest rampant polyphyly across the *C. carnea* complex (Haruyama, Naka, et al., 2008; Lourenço et al., 2006; Price et al., 2015). General

patterns of species placement in mitochondrial phylogenies suggest geographic clades with multiple evolutionary changes in the morphological, ecological, and acoustic characteristics that are known to vary among species. Apparent polyphyly of mitochondrial haplotypes may be the result of incomplete lineage sorting or mitochondrial capture rather than a reflection of the complete evolutionary history of the group. Sanger sequencing of a small number of nuclear genes has proved insufficient to resolve relationships within the *carnea*-group (Haruyama, Mochizuki, et al., 2008). As in other rapid radiations, a large number of loci from across the nuclear genomes will likely be necessary to resolve the true evolutionary history of the *carnea*-group (Nadeau et al., 2013; Wagner et al., 2013).

The lack of reliable phylogenetic resources has contributed to taxonomic instability and species delimitation troubles in the *carnea*-group (Henry et al., 2013; Price et al., 2015; Thierry et al., 2011). Scientists studying these insects have applied different species boundaries and naming conventions, making literature across research groups difficult to interpret. Confusion and controversy over species delimitation may have hindered ongoing study of this once-emerging speciation study system.

Lacewings in the genus *Chrysoperla* are among the most important biological control agents used against insect agricultural pests (Pappas et al., 2011). *Chrysoperla carnea*-group insects sold for biological control are often incorrectly identified (Henry & Wells, 2007). Researchers and biological control companies agree that an understanding of the systematics of the *carnea*-group is needed to inform decision-making regarding their use in integrated pest management strategies (Henry & Wells, 2007; Pappas et al., 2011; Tauber et al., 2000).

Proposed diversification processes

With current data we cannot evaluate the multiple existing explanations for the rapid

diversification in the *carnea*-group. Tauber and Tauber (reviewed in Tauber & Tauber, 1981) have proposed a process of speciation through life cycle switches and resultant breeding asynchrony in sympatry. The species used as the basis of their sympatric speciation model, *Chrysoperla plorabunda* and *Chrysoperla downesi*, are not supported as sister species in recent mitochondrial phylogenies (Price et al. 2015). Thierry et al. (2011) have proposed that competition between *carnea*-group lacewings for resources has led to ecological specialization on habitat types and that songs evolved later as secondary barriers to hybridization. The two populations that seemed to demonstrate ecological character displacement, called 'C. affinis a' (song-delineated species *C. agilis*) and 'C. affinis s*'* (song-delineated species *C. carnea*), seem to not be sister lineages based upon mitochondrial phylogenies, and have distinct mating songs. Henry et al. (2013) proposed speciation through sexual selection, facilitated by habitat heterogeneity and the geographic isolation of small populations. However, Noh and Henry (2015) found only weak support for the idea that sexual selection is acting on the same song elements that are used for species recognition, and thus there is limited evidence supporting a continuous process beginning with female choice that leads to speciation through divergent species recognition mechanisms. It is unclear which of the above explanations is closest to the truth; perhaps all of these factors played a role in the diversification of this clade.

Plausible evolutionary scenarios

A. Groups with shared ecological niches

Suites of potentially adaptive traits appear together in multiple *carnea*-group species with similar ecological niches on different continents. This morphological and acoustic similarity could be explained by descent from a single ancestor sharing those traits and the same niche. For example, conifer-associated *carnea*-group species have a dark green ground color and retain that

color through the winter, while other species occupying deciduous vegetation have a lighter green ground coloration and turn shades of brown to match their habitat with the seasons (Duelli et al., 2014). Conifer-associated species also tend to perform relatively long courtship signals composed of many short bursts of sounds before waiting for a response from a duetting partner. This pattern could be explained by a single origin of this suite of traits (i.e., conifer-feeding, dark, unchanging coloration, and long courtship signal) in the common ancestor of a clade of closely related species. In that case the occurrence of conifer species on multiple continents could be explained by the strong dispersal ability of these species (Duelli, 1984).

B. Geographic groups

Patterns in mitochondrial DNA phylogenies suggest that lacewings found on the same continent are most closely related (Price et al., 2015). However, this is not supported by the taxonomic distribution of morphological or acoustic traits, and is unexpected for organisms with strong dispersal abilities. This scenario implies multiple convergent evolutionary change in numerous traits, which could possibly be due to similar evolutionary pressures acting on a shared genetic background.

C. A history of introgression

Among the recently diverged members of the *Chrysoperla carnea*-group, mitochondrial analyses (Price et al., 2015), behavioral song response data (Noh & Henry, 2010), and frequent syntopy suggest that current or historical introgression might be possible. Patterns of complex traits observed in the *carnea*-group might be explained by introgression, for example in the dry-associated ‘mohave’ ecotype. Two undescribed species, ‘Chrysoperla mohave 1’ and ‘Chrysoperla mohave 2’, have distinct pale body coloration, dark cross veins, red facial and abdominal markings, specific diapause emergence cues, and a set of habitat associations unlike

other members of the *carnea*-group. This complex suite of traits seems closely associated with dry climates and less green foliage. The two 'mohave' types each share a vibrational mating song with another parapatric species from western North America, either *Chrysoperla johnsoni* or *Chrysoperla downesi*. Adaptive introgression between the two 'mohave' types could explain the presence of this suite of desiccation-related traits in the two song groups.

Goals

With genome-wide genetic markers, we explored the evolutionary history of the rapidly diversifying *Chrysoperla carnea* group. We reconstructed species relationships within the group using phylogenetic methods and identified genetic groups using population clustering methods. We additionally tested for evidence of introgression in the diversification history of the clade. Specifically, we explored support for three evolutionary scenarios outlined above, i.e., (A) groups with shared ecological niches, (B) geographic groups, and (C) a history of introgression. This work identified the patterns of diversification in the recent and rapid radiation of *Chrysoperla carnea*-group species and serves as a road map for future work in this unique system for the genomic study of speciation.

METHODS:

Taxon sampling

Samples were selected from Charles Henry's collection of preserved (dry frozen or frozen in EtOH) lacewings. Specimens were identified before preservation by Charles Henry or Peter Duelli based upon song and morphology. Two to seven individuals of each described species and ecotype were selected for sequencing, with individuals selected from as many distinct regions of their range as possible (**Table 1**).

RAD sequencing, locus construction, and quality control

Samples from each species and ecotype were spread evenly across three RAD sequencing libraries. The libraries were constructed with the SbfI restriction enzyme at the Institute for System Genomics at the University of Connecticut, and each pool was sequenced on two lanes of an Illumina HiSeq 4000. The reads were demultiplexed using process_radtags in the STACKS 2.0 pipeline (Rochette et al., 2019) with no quality control filters enabled. Illumina adaptors and low-quality regions were trimmed with Trimmomatic (Bolger et al., 2014). Quality-controlled reads were aligned to the *Chrysoperla carnea* reference genome (SUB6529655) with BWA (Li & Durbin, 2009), and the SAM file outputs were converted to sorted bam files using SAMtools (Li, 2011). Loci were constructed with the default setting of gstacks in the STACKS 2.0 pipeline (Rochette et al., 2019).

Loci potentially from contaminant sequence were identified and removed based upon a BLAST (Camacho et al., 2009) search of the loci catalogue against the NCBI nucleotide database. Low complexity loci with average GC content less than 10% and loci average GC content higher than expected from these organisms, greater than 50%,, were removed from the catalogue. Loci present in fewer than four individuals were also removed.

Single nucleotide polymorphisms (SNPs) in the catalogue loci were output with the populations module of STACKS 2.0 (Rochette et al., 2019). SNPs with a mean depth of greater than 12,000, or minor allele count of less than 2, were filtered with VCFtools (Danecek et al., 2011). *Chrysoperla carnea*-group individuals with sequence at fewer than 1000 sites or with less than 10% of reads passing quality control and mapping to the reference genome were removed from the analysis. The full SNP datasets contained 166,630 variable sites for 126 individuals across all 22 *carnea*-group species and four outgroup species.

Phylogenetic reconstruction

The maximum likelihood phylogeny was generated with IQtree (Minh et al., 2020), from a concatenated SNP matrix of the 126 individuals. A transversion model with empirical base frequencies and three rate categories (TVM+F+R3), was selected based upon BIC with ModelFinder (Kalyaanamoorthy et al., 2017), and the proportion of invariant sites was fixed at zero. One hundred maximum likelihood non-parametric bootstrap replicates were calculated in IQtree to measure the support for tree edges. Site concordance factor was calculated across one thousand quartets in IQtree to assess conflict in the data set.

Coalescent-based species trees were produced by quartet sampling with SVDquartets (Chifman & Kubatko, 2014) as implemented in PAUP* (Swofford, 2003). For this analysis, each monophyletic cluster of individuals hypothesized to be conspecific (based upon mating signal identification) was treated as a species. Three individuals resolved outside of monophyletic species clusters, 171 adamsi, 134 plorabunda, and 096 carnea were removed, and as *C. johnsoni* did not form a single monophyletic cluster, it was separated into two monophyletic groups 'johnsoni A' and 'johnsoni B'. All quartets in the data set were examined for species-tree topology construction. One hundred bootstrap replicates, each examining 100,000 random quartets representing 1.37% of the total quartets, were calculated to assess support for the topology obtained from all quartets. Branch lengths for the fixed species-tree topology were estimated with PAUP* (Swofford, 2003). A general time reversible model with empirical base frequencies, gamma shape rate parameter, and the proportion of invariant sites fixed at zero was used for branch length estimation. All phylogenetic trees were plotted using FigTree (Rambaut, 2012). Summarized geographic and ecological associations for each species were displayed on each tip along with a representative courtship song oscillogram. Oscillograms for each species were selected from the published literature (Henry, 1992; Henry et al., 2013, 2018, 2019).

Population genomics

Population genetic structure was assessed on an unlinked loci set, one randomly selected SNP from each RAD locus, with Bayesian clustering implemented in STRUCTURE (Pritchard et al., 2000). All individuals in the dataset were clustered into 1–22 populations (k) in three independent runs for each k value over 50,000 iterations with the first 5,000 discarded as burn-in. The mean probability of the data for each value of K, L(K), and the change in log probability of the data for successive k values, Delta K (Evanno et al., 2005), were calculated and plotted with STRUCTURE HARVESTER (Earl & vonHoldt, 2012). For the values of k with high L(K) and Delta K the three runs were summarized using CLUMPP (Jakobsson & Rosenberg, 2007) using the greedy algorithm or large K greedy algorithm executed across all possible input orders. The population assignments calculated with CLUMPP were plotted using the R package strataG (Archer et al., 2017).

Introgression analysis

Hybridization was identified based on deviations from expected site pattern frequency as calculated by the Python script HyDe (Blischak et al., 2018). Only monophyletic species clusters were used in this analysis. The individuals removed from species-tree estimation were also removed from introgression tests. The three-taxon test was run on all possible combinations of ingroup triplets for a total of 5,313 tests. The clade of all non-*carnea* group samples was used as the outgroup for all tests.

Results

Phylogenetic reconstruction

Bifurcating phylogenetic trees were inferred using maximum likelihood (**Fig 1**) and the species-tree method SVDquartets (**Fig 2**). The maximum likelihood approach resolved the

majority of song species as monophyletic and both approaches resolved primarily geographic groupings above the species level. Conflicting signal or lack of signal resulted in low support values for some clades, especially in the non-conifer associated North American species and for some deeper branches in the tree. Outside of that group, a single *C. carnea* individual from Iran, 096 carnea, is nested within with *C. heidarii,* possibly due to sample mis-identification, mix-up, or overlooked data quality problems.

Species clustering

Individuals were clustered into five, nine, or 16 genetic groups (**Fig 3**). When clustered into five groups, clear clusters emerged for each of the four major geographic groups. The fifth cluster contains primarily the species *C. carnea*. Many individuals from European and West Asian species and all of the North American conifer-associated species have some probability of assignment to the fifth cluster containing *C. carnea*. With larger numbers of clusters more individual species groups emerged, but also the conflicting signals in the dataset became clearer. For example, *C. lucasina* and *C. heidarii* individuals never seemed to cluster clearly into any group; this may be due to the small number of individuals representing these species or high levels of missing data for these samples. Other species, including *C. zastrowi sillemi*, *C. carnea*, and 'C. mohave 2,' each comprised multiple individuals with a high probability of assignment to more than one cluster at multiple values of K.

Introgression analysis

Of the 5,313 potential species triplets in the dataset, 149 showed significant evidence of introgression **(Table 2).** One of the significant triplets included the two 'mohave' types and *C. johnsoni* B (the species with which 'C. mohave 1' shares a song), showed significant evidence of introgression. *C. johnsoni* A was involved in more significant introgression tests than any other

species in the analysis, a total of 26 tests. Seven of those significant triplets involved *C. johnsoni* A and *C. johnsoni* B. This suggests that introgression may underlie the two distinct phylogenetic clusters this species forms in the maximum likelihood and species phylogenies.

Discussion

In this study, we have begun to unravel the evolutionary history of the *carnea*-group, a task made challenging by the recent and rapid diversification of the group and its history of introgression. Despite these challenges we now have the clearest picture yet of the evolutionary history of the *Chrysoperla carnea* species group. We find that the deepest phylogenetic structure groups species by geography, with four major groups: the European and West Asian species, the North American species, the East Asian species, and the southerly species. Each forms a distinct group in all analyses. These geographic clades with replicated radiation on each continent are unexpected given the formidable dispersal abilities of these insects (Duelli, 1984), but align with prior mitochondrial DNA analyses (Price et al., 2015). Rather than species that share morphological, ecological, or behavioral traits being closely related, the overall topology indicates multiple evolutionary origins of similar habitat associations, body coloration, and mating song features. For example, among the conifer associates *C. downesi*, *C. calocedrii*, *C. duellii*, and *C. mediterranea*, only the two North American forms *C. calocedrii* and *C. downesi* are closely related. The general patterns of diversification in the group suggest convergent or parallel evolution of multiple traits, perhaps underlaid by a replicated adaptive radiation and ecological niche-filling on each continent. This replicated evolution of ecomorphs is similar to the replicated divergence found in some island taxa like the *Anolis* lizards and Hawaiian stick spiders (Gillespie et al., 2018; Losos, 2011), but unexpected between continents and for organisms with strong dispersal capabilities (Losos, 2010).

Overall, we find strong support for monophyletic grouping or genetic clustering of most of the behaviorally, ecologically, and morphologically delineated species, with the notable exception of the non-conifer associated species and ecotypes in North America (*C. plorabunda*, *C. adamsi*, *C. johnsoni*, 'C. mohave 2', and 'C. mohave 1'), potentially due to very recent divergence. Morphological traits, to some extent, seem to correspond to groupings above the species level; for example, *C. lucasina* and *C. heidarii* cluster together in both phylogenies and share a dark stripe on the pleuron of the second abdominal segment, a feature not observed in other species in the group (Henry et al., 2014).

The species with similar song characteristics are not closely related. This finding supports the hypothesis of multiple parallel speciation and the evolution of identical mating song on different continents in the *Chrysoperla carnea*-group (Henry et al., 1999, 2012, 2014, 2019). This convergence has been proposed to be caused by a random process that can generate functionally similar song types in geographically isolated areas through simple genetic control and competitive reproductive exclusion due to limited acoustic niche space. Alternatively, this could result from adaptation to different transmission qualities of plant substrate, but experimental evidence does not support that notion (Henry & Wells, 2004).

With this large dataset of genome-wide nuclear markers, we find that it is possible to distinguish most of the *carnea*-group species. A large number of loci from across the genome have been necessary to differentiate species in other rapid radiations (Wagner et al., 2013). Future development and testing of markers for identification based upon this dataset could allow other scientists to quickly and accurately identify many of these cryptic species. This will be especially useful for commercial insectaries and scientists working to improve the application of these generalist predators for biological control of crop pests. For example, identification of

insectary stocks to species is critical for targeting the inundative release of *carnea*-group species to the correct crop type and best ecological context.

Prior discordance observed in the mitochondrial dataset was expected to be due to incomplete lineage sorting or introgressive hybridization. It is therefore not surprising we found support for these very recently diverged song lineages persisting with a history of gene flow and introgression. There is strong evidence in particular that introgression has played a role in the diversification of *Chrysoperla johnsoni*, as supported by the significant tests for introgression involving those groups, placement of this clade into two clades in the phylogenetic trees, and patterns of population clustering. We furthermore predicted that introgression may have played a role in the acquisition of either song or desiccation tolerance traits in the 'mohave' ecotypes, and we found some support for this idea in the significant test for introgression involving these species. These findings of introgression suggest that potentially some regions of the genome may be free to introgress, while those regions related to barrier behavioral and ecological traits have remained distinct, as has been found in *Heliconius* butterflies (Martin et al., 2013). Alternatively, the history of introgression may have at least in part fueled radiation.

Overall, we significantly advance the understanding of the evolutionary history of *Chrysoperla carnea*-group lacewings. We show for the first time that molecular evidence supports the species known from song, morphology, and ecology. We find a repeated radiation of ecomorphs in North American and Eurasia. Furthermore, we have come to appreciate the role of introgression in the divergence of these lineages. Going forward, scientists can begin to synthesize forty years of descriptive work on these taxa by linking this genetic data with morphological, behavioral, and ecological data.

Tables and figures

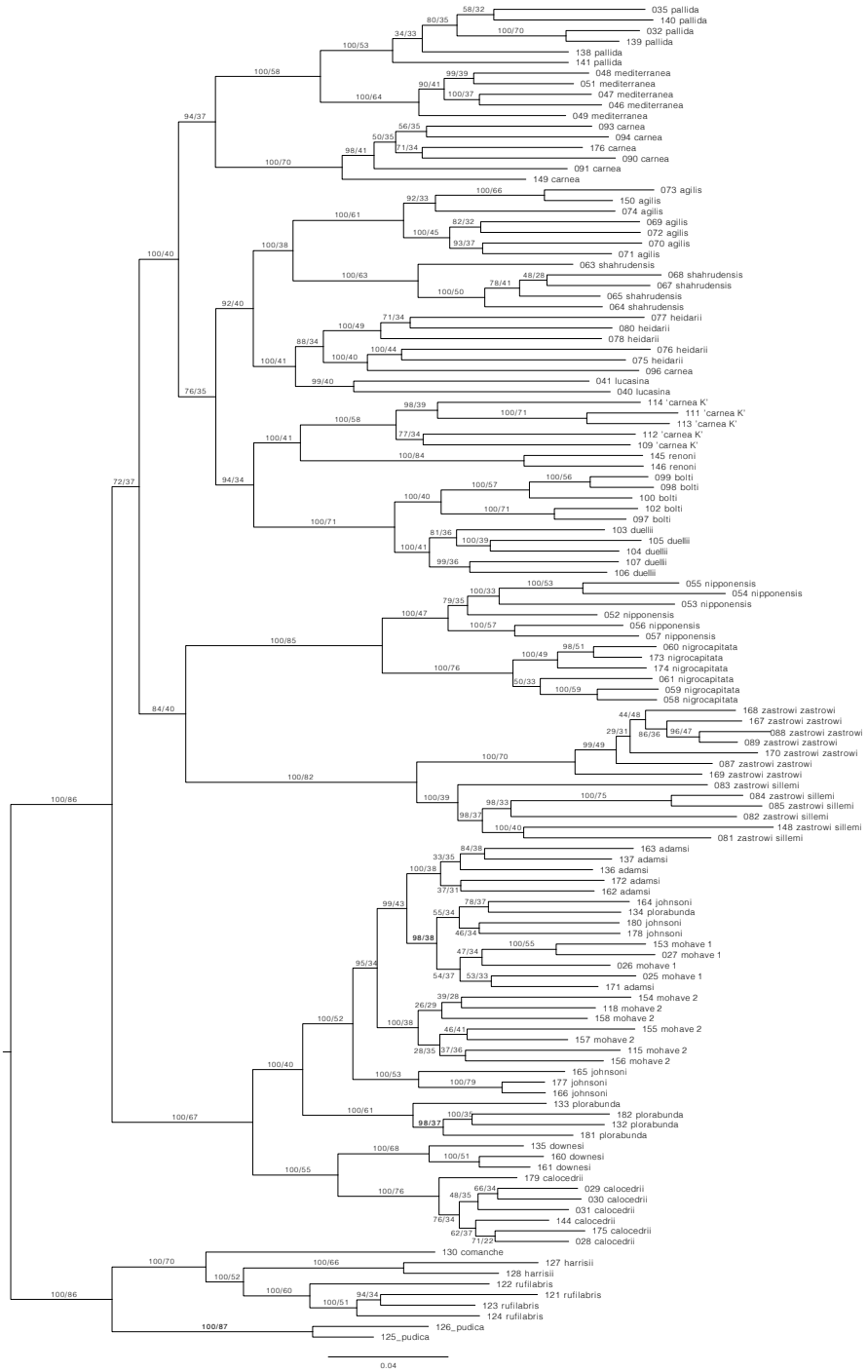

Figure 1. Maximum likelihood phylogeny inferred for the *Chrysoperla carnea*-group of species with IQtree using the TVM+F+R3 model of evolution. Bootstrap replicate values are displayed first on branches followed by site concordance factors.

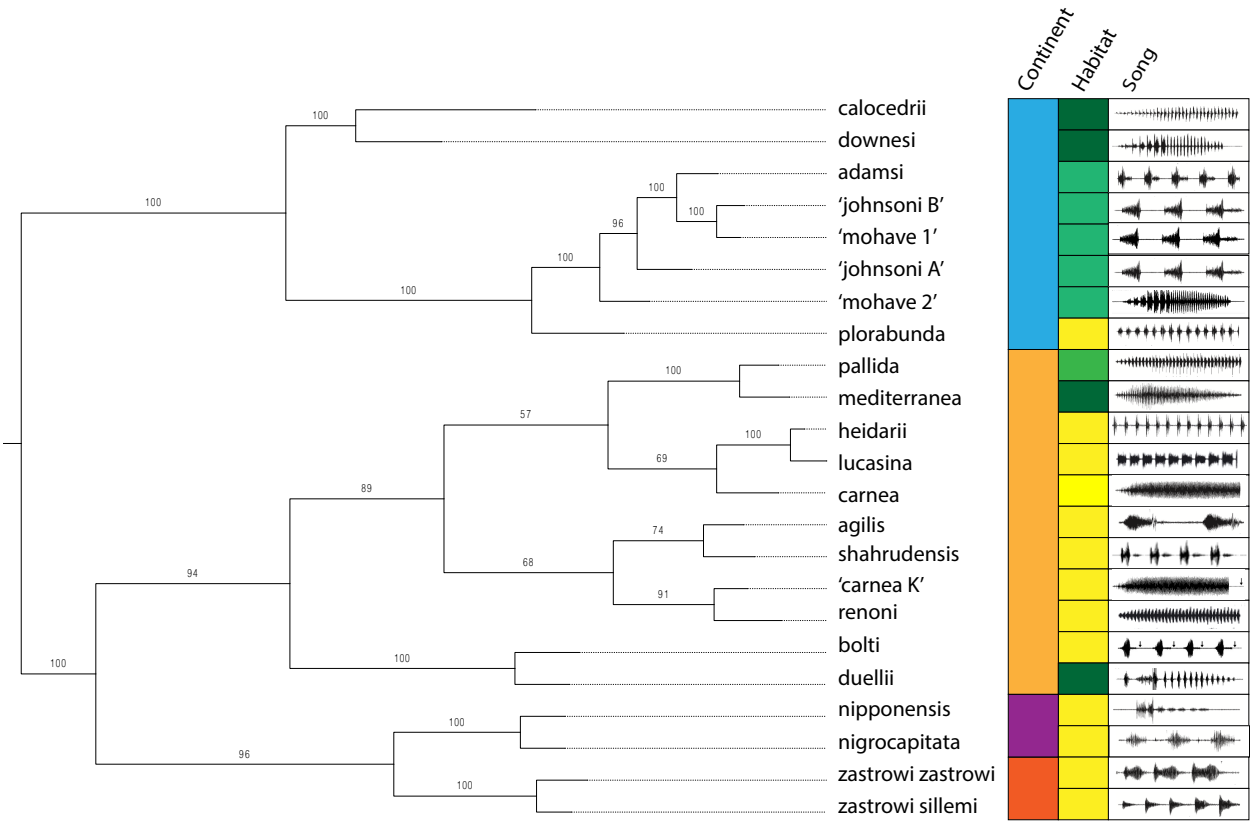

Figure 2. Species tree of the *Chrysoperla carnea*-group calculated with SVDQuartets. Bootstrap replicate values are displayed on the branches. General geographic area for each species is indicated by color: blue for North America, orange for Europe and West Asia, purple for East Asia, and red for southerly species. Habitat association is indicated by color: dark green for evergreen conifer-associated species, light green for species with mixed habitat association, and yellow for field or meadow associated species. Representative oscillograms of courtship signals are taken from the primary literature (see text for references).

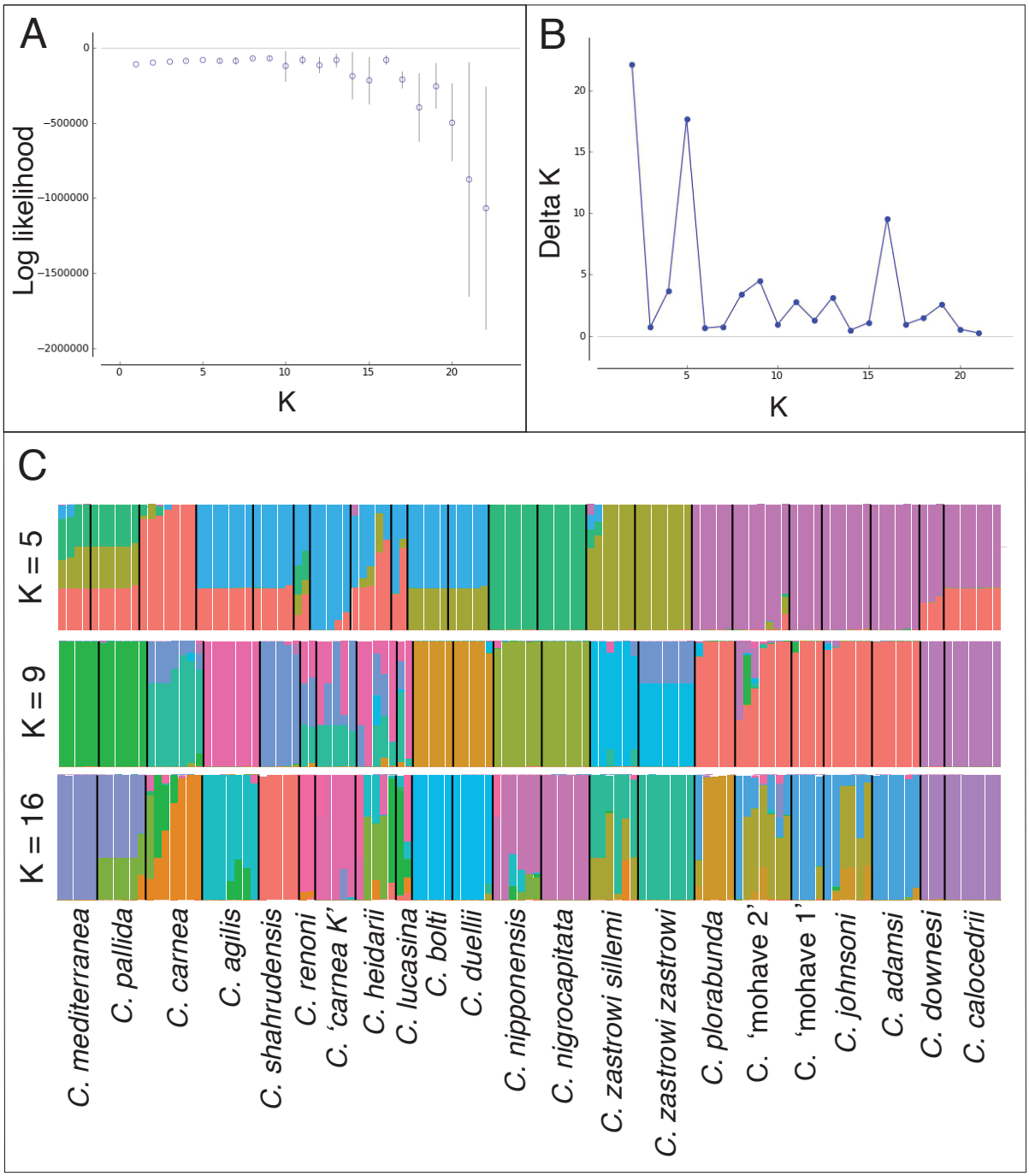

Figure 3. Genetic cluster inference with STRUCTURE. A. Mean log likelihood of the data for each number of clusters (K). B. Increase in likelihood (Delta K) for each successive value of K. C. Group assignment probability for each individual in 5, 9, and 16 clusters.

References

Archer, F. I., Adams, P. E., & Schneiders, B. B. (2017). STRATAG: An R package for manipulating, summarizing and analysing population genetic data. *Molecular Ecology Resources*, *17*(1), 5–11. https://doi.org/10.1111/1755-0998.12559

Blischak, P. D., Chifman, J., Wolfe, A. D., & Kubatko, L. S. (2018). HyDe: A python package for genome-scale hybridization detection. *Systematic Biology*, *67*(5), 821–829. https://doi.org/10.1093/sysbio/syy023

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.

Brooks, S. J. (1994). A taxonomic review of the common green lacewing genus Chrysoperla (Neuroptera: Chrysopidae). *Bulletin of the British Museum of Natural History (Entomology)*, *63*(2), 137–210.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST : architecture and applications. *BMC Bioinformatics*, *10*(1), 421.

Chifman, J., & Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model. *Bioinformatics*, *30*(23), 3317–3324. https://doi.org/10.1093/bioinformatics/btu530

Cook, L. M. (2008). Species richness in Madeiran land snails, and its causes. *Journal of Biogeography*, *35*(4), 647–653. JSTOR.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Duelli, P. (1984). Dispersal and opposition strategies in *Chrysoperla carnea*. *Progress in World's Neuropterology*, 133–145.

Duelli, P., Johnson, J. B., Waldburger, M., & Henry, C. S. (2014). A new look at adaptive body coloration and color change in "common green lacewings" of the genus *Chrysoperla* (Neuroptera: Chrysopidae). *Annals of the Entomological Society of America*, *107*(2), 382–388.

Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*(2), 359–361. https://doi.org/10.1007/s12686-011-9548-7

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, *14*, 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Gillespie, R. G., Benjamin, S. P., Brewer, M. S., Rivera, M. A. J., & Roderick, G. K. (2018). Repeated diversification of ecomorphs in Hawaiian stick spiders. *Current Biology*, *28*(6), 941-947.e3. https://doi.org/10.1016/j.cub.2018.01.083

Grant, P. R., & Grant, B. R. (2011). *How and Why Species Multiply: The Radiation of Darwin's Finches*. Princeton University Press.

Harrison, R. G., & Larson, E. L. (2014). Hybridization, Introgression, and the Nature of Species Boundaries. *Journal of Heredity*, *105*(S1), 795–809. https://doi.org/10.1093/jhered/esu033

Haruyama, N., Mochizuki, A., Duelli, P., Naka, H., & Nomura, M. (2008). Green lacewing phylogeny, based on three nuclear genes (Chrysopidae, Neuroptera). *Systematic Entomology*, *33*(2), 275–288. https://doi.org/10.1111/j.1365-3113.2008.00418.x

Haruyama, N., Naka, H., Mochizuki, A., & Nomura, M. (2008). Mitochondrial phylogeny of cryptic species of the lacewing *Chrysoperla nipponensis* (Neuroptera: Chrysopidae) in Japan. *Annals of the Entomological Society of America*, *101*(6), 971–977. https://doi.org/10.1603/0013-8746-101.6.971

Henry, C. (1992). *Chrysoperla mohave* (Banks) Neuroptera: Chrysopidae: Two familiar species in an unexpected disguise. *Psyche: A Journal of Entomology*, *99*. https://doi.org/10.1155/1992/16818

Henry, C. S., Brooks, S. J., Duelli, P., Johnson, J. B., Wells, M. L. M., & Mochizuki, A. (2012). Parallel evolution in courtship songs of North American and European green lacewings (Neuroptera: Chrysopidae). *Biological Journal of the Linnean Society*, *105*(4), 776–796. https://doi.org/10.1111/j.1095-8312.2011.01845.x

Henry, C. S., Brooks, S. J., Duelli, P., Johnson, J. B., Wells, M. M., & Mochizuki, A. (2013). Obligatory duetting behaviour in the Chrysoperla carnea ‐ group of cryptic species (Neuroptera: Chrysopidae): its role in shaping evolutionary history. *Biological Reviews*, *88*(4), 787–808.

Henry, C. S., Brooks, S. J., Johnson, J. B., Mochizuki, A., & Duelli, P. (2014). A new cryptic species of the Chrysoperla carnea group (Neuroptera: Chrysopidae) from western Asia: parallel speciation without ecological adaptation. *Systematic Entomology*, *39*(2), 380–393. https://doi.org/10.1111/syen.12061

Henry, C. S., Brooks, S. J., Johnson, J. B., Mochizuki, A., Mirmoayedi, A., & Duelli, P. (2018). Distinctive but functionally convergent song phenotypes characterize two new allopatric species of the *Chrysoperla carnea* -group in Asia, *Chrysoperla shahrudensis* sp. nov. and

Chrysoperla bolti sp. nov. (Neuroptera: Chrysopidae). *Journal of Natural History*, *52*(25–26), 1603–1635. https://doi.org/10.1080/00222933.2018.1478011

Henry, C. S., Taylor, K. L., & Johnson, J. B. (2019). A new lacewing species of the Chrysoperla carnea species-group from central Asia associated with conifers (Neuroptera: Chrysopidae). *Journal of Natural History*, *53*(21–22), 1277–1300. https://doi.org/10.1080/00222933.2019.1644385

Henry, C. S., & Wells, M. L. M. (2004). Adaptation or random change? The evolutionary response of songs to substrate properties in lacewings (Neuroptera: Chrysopidae: Chrysoperla). *Animal Behaviour*, *68*(4), 879–895.

Henry, C. S., Wells, M. L. M., & Holsinger, K. E. (2002). The inheritance of mating songs in two cryptic, sibling lacewing species (Neuroptera: Chrysopidae: *Chrysoperla*). In *Genetics of Mate Choice: From Sexual Selection to Sexual Isolation* (Vol. 1–Book, Section, pp. 269–289). Springer.

Henry, C. S., Wells, M. L. M., & Simon, C. M. (1999). Convergent evolution of courtship songs among cryptic species of the *carnea* group of green lacewings (Neuroptera: Chrysopidae: *Chrysoperla*). *Evolution*, *53*(4), 1165–1179. JSTOR. https://doi.org/10.2307/2640820

Henry, C. S., & Wells, M. M. (2007). Can what we don't know about lacewing systematics hurt us? A cautionary tale about mass rearing and release of "*Chrysoperla carnea*" (Neuroptera: Chrysopidae). *American Entomologist*, *53*(1), 42–47. https://doi.org/10.1093/ae/53.1.42

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population

structure. *Bioinformatics*, *23*(14), 1801–1806. https://doi.org/10.1093/bioinformatics/btm233

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Kozak, K. H., Weisrock, D. W., & Larson, A. (2006). Rapid lineage accumulation in a non-adaptive radiation: phylogenetic analysis of diversification rates in eastern North American woodland salamanders (Plethodontidae: *Plethodon*). *Proceedings of the Royal Society B: Biological Sciences*, *273*(1586), 539–546. https://doi.org/10.1098/rspb.2005.3326

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Losos, J. B. (2010). Adaptive Radiation, Ecological Opportunity, and Evolutionary Determinism: American Society of Naturalists E. O. Wilson Award Address. *The American Naturalist*, *175*(6), 623–639. https://doi.org/10.1086/652433

Losos, J. B. (2011). *Lizards in an evolutionary tree: ecology and adaptive radiation of Anoles*. Univ of California Press.

Lourenço, P., Brito, C., Backeljau, T., Thierry, D., & Ventura, M. A. (2006). Molecular systematics of the *Chrysoperla carnea* group (Neuroptera: Chrysopidae) in Europe.

Journal of Zoological Systematics and Evolutionary Research, *44*(2), 180–184. https://doi.org/10.1111/j.1439-0469.2006.00352.x

Macleod, E. G. (1967). Experimental induction and elimination of adult diapause and autumnal coloration in Chrysopa carnea (Neuroptera). *Journal of Insect Physiology*, *13*(9), 1343–1349. https://doi.org/10.1016/0022-1910(67)90136-9

Malinsky, M., Svardal, H., Tyers, A. M., Miska, E. A., Genner, M. J., Turner, G. F., & Durbin, R. (2018). Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecology & Evolution*, *2*(12), 1940–1955. https://doi.org/10.1038/s41559-018-0717-x

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, *20*(5), 229–237. https://doi.org/10.1016/j.tree.2005.02.010

Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., Blaxter, M., Manica, A., Mallet, J., & Jiggins, C. D. (2013). Genome-wide evidence for speciation with gene flow in Heliconius butterflies. *Genome Research*, *23*(11), 1817–1828. https://doi.org/10.1101/gr.159426.113

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msaa015

Nadeau, N. J., Martin, S. H., Kozak, K. M., Salazar, C., Dasmahapatra, K. K., Davey, J. W., Baxter, S. W., Blaxter, M. L., Mallet, J., & Jiggins, C. D. (2013). Genome-wide patterns of divergence and gene flow across a butterfly radiation. *Molecular Ecology*, *22*(3), 814–826. https://doi.org/10.1111/j.1365-294X.2012.05730.x

Noh, S., & Henry, C. S. (2010). Sexually monomorphic mating preferences contribute to premating isolation based on song in european green lacewings. *Evolution*, *64*(1), 261–270. https://doi.org/10.1111/j.1558-5646.2009.00828.x

Papadopoulou, A., Anastasiou, I., & Vogler, A. P. (2010). Revisiting the insect mitochondrial molecular clock: the mid-aegean trench calibration. *Molecular Biology and Evolution*, *27*(7), 1659–1672. https://doi.org/10.1093/molbev/msq051

Pappas, M., Broufas, G., & Koveos, D. (2011). Chrysopid predators and their role in biological control. *Journal of Entomology*, *8*(3), 301–326.

Pardo-Diaz, C., Salazar, C., Baxter, S. W., Merot, C., Figueiredo-Ready, W., Joron, M., McMillan, W. O., & Jiggins, C. D. (2012). Adaptive introgression across species boundaries in *Heliconius* Butterflies. *PLoS Genetics*, *8*(6), e1002752. https://doi.org/10.1371/journal.pgen.1002752

Price, B. W., Henry, C. S., Hall, A. C., Mochizuki, A., Duelli, P., & Brooks, S. J. (2015). Singing from the grave: DNA from a 180 year old type specimen confirms the identity of *Chrysoperla carnea* (Stephens). *PLoS ONE*, *10*(4), e0121127. https://doi.org/10.1371/journal.pone.0121127

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959.

Rambaut, A. (2012). *FigTree v1. 4*. http://tree.bio.ed.ac.uk/software/figtree/

Rochette, N. C., Rivera‐Colón, A. G., & Catchen, J. M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, *28*(21), 4737–4754. https://doi.org/10.1111/mec.15253

Rundell, R. J., & Price, T. D. (2009). Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. *Trends in Ecology & Evolution*, *24*(7), 394–399. https://doi.org/10.1016/j.tree.2009.02.007

Schluter, D. (2001). Ecology and the origin of species. *Trends in Ecology & Evolution*, *16*(7), 372–380. https://doi.org/10.1016/S0169-5347(01)02198-X

Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends in Ecology & Evolution*, *19*(4), 198–207. https://doi.org/10.1016/j.tree.2004.01.003

Seehausen, O. (2006). African cichlid fish: a model system in adaptive radiation research. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1597), 1987–1998. https://doi.org/10.1098/rspb.2006.3539

Swofford, D. L. (2003). *PAUP*: Phylogenetic Analysis Using Parsimony (and other methods) 4.0.b5*.

Tauber, C A, & Tauber, M. J. (1981). Insect Seasonal Cycles: Genetics and Evolution. *Annual Review of Ecology and Systematics*, *12*(1), 281–308. https://doi.org/10.1146/annurev.es.12.110181.001433

Tauber, Catherine A., & Tauber, M. J. (1977). A genetic model for sympatric speciation through habitat diversification and seasonal isolation. *Nature*, *268*(5622), 702.

Tauber, Catherine A., Tauber, M. J., & Nechols, J. R. (1977). Two genes control seasonal isolation in sibling species. *Science*, *197*(4303), 592–593. https://doi.org/10.1126/science.197.4303.592

Tauber, M. J., & Tauber, C. A. (1973). Nutritional and photoperiodic control of the seasonal reproductive cycle in *Chrysopa mohave* (Neuroptera). *Journal of Insect Physiology*, *19*(4), 729–736. https://doi.org/10.1016/0022-1910(73)90145-5

Tauber, M. J., Tauber, C. A., Daane, K. M., & Hagen, K. S. (2000). Commercialization of predators: recent lessons from green lacewings (Neuroptera: Chrysopidae: Chrosoperla). *American Entomologist*, *46*(1), 26–38. https://doi.org/10.1093/ae/46.1.26

Thierry, D., Canard, M., Deutsch, B., Ventura, M. A., Lourenço, P., & Lodé, T. (2011). Ecological character displacement in competing common green lacewings in Europe: a route to speciation? *Biological Journal of the Linnean Society*, *102*(2), 292–300. https://doi.org/10.1111/j.1095-8312.2010.01569.x

Thierry, D., Cloupeau, R., Jarry, M., & Canard, M. (2010). Discrimination of the West-Palaearctic *Chrysoperla* Steinmann species of the *carnea* Stephens group by means of claw morphology (Neuroptera, Chrysopidae). *Acta Zoologica Fennica*, *209*, 255–262.

Wagner, C. E., Keller, I., Wittwer, S., Selz, O. M., Mwaiko, S., Greuter, L., Sivasundar, A., & Seehausen, O. (2013). Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, *22*(3), 787–798. https://doi.org/10.1111/mec.12023

Candidate genomic region for species-specific mating song in lacewings

Katherine Taylor, Elizabeth Wade, Marta Wells, Charles Henry

Introduction

Identifying the genomic underpinnings of traits important to reproductive isolation has the potential to inform our understanding of the process of speciation. Determining the number, relative positions, and effect size of loci underlying these traits is fundamental to understanding the dynamics of the speciation process (Coyne, 1992; Templeton, 1981). With access to abundant and affordable sequencing data and genomic resources for many organisms, we are beginning to answer these long-standing questions (Seehausen et al., 2014; Wu & Ting, 2004). Researchers have begun to identify the genetic underpinnings of traits important to reproductive isolation in many different organisms (Blankers et al., 2019; Kronforst et al., 2006; Nosil & Schluter, 2011).

Species-specific mating signals often facilitate divergence and speciation in rapid radiations (Wilkins et al., 2013). *Drosophila* and crickets in the genus *Laupala* are two insect groups with species-specific acoustic signals performed by males prior to mating. The mechanism of song production is quite different in these groups. *Drosophila* sing with wing vibration while *Laupala* use stridulatory organs to produce sound. Both types of acoustic signals are critical to reproductive isolation between species. The genetic basis of insect mating signals has been most extensively explored in *Drosophila* and *Laupala*. In both genera, mating song phenotypes are controlled by many genomic loci (Mullen & Shaw, 2014).

In *Laupala*, mating signals have been associated with between seven and nine quantitative trait loci (Blankers et al., 2019). Genetic linkage between signal phenotype and signal preference has been demonstrated in multiple species pairs (Shaw & Lesnick, 2009; Wiley

et al., 2012; Xu & Shaw, 2019). Tight genetic linkage between these traits may facilitate rapid speciation in the *Laupala* group (Xu & Shaw, 2019).

In *Drosophila melanogaster*, many genes have been identified that are necessary for the production of typical song phenotype or for the perception of that phenotype (Gleason, 2005; Greenspan & Ferveur, 2000). These genes have a wide variety of other functions, including sex determination, control of circadian rhythms, wing function, and sensory organ function (Gleason, 2005; Greenspan & Ferveur, 2000). However, genes and mutations known to modify mating signals in the laboratory are not necessarily the genes that drive species-specific difference in songs (Gleason, 2005).

One approach to identify the genes that underlie a trait of interest that varies between lineages is the association of genomic regions with phenotypes in hybrid cross progeny. Those regions, or quantitative trait loci (QTLs), can be detected by examining the correlation of phenotypes to markers throughout the genome. With this approach, only regions related to the phenotype of interest that vary between the lines crossed will be identified. However, fine mapping of traits to small genomic windows, which is necessary to identify the actual evolutionary changes, is dependent on the number of individuals in the cross, recombination frequency, marker density, and strength of the effect of the loci (Broman, 2001; Mackay et al., 2009).

An alternate or complementary approach to identify regions containing traits important to reproductive isolation is genome scans for the signal of divergence (Seehausen et al., 2014). Calculating divergence statistics across the genome can reveal patterns related to selection or a lack of introgression near traits of interest. Divergence scans between populations have revealed great variance in divergence across the genome (Nosil & Feder, 2012). Highly diverged regions

may correspond to regions of the genome not free to introgress due to strong selection. In many cases it has been found that regions of high divergence are not related to admixture but rather to other demographic processes (Cruickshank & Hahn, 2014). However, between those lineages with current admixture we would still expect regions of the genome related to barrier loci to show greater differentiation (Campbell et al., 2018).

Species in the *Chrysoperla carnea*-group of green lacewings are reproductively isolated by species-specific premating duets (Henry et al., 2013). Species in this morphologically cryptic group vary primarily in their mating signals. Lacewings produce vibrational courtship signals by vigorous but highly controlled shaking of the thorax and abdomen. Male and female lacewings in the *carnea*-group produce nearly identical signals that are exchanged in an intricate duet prior to copulation. These signals are thought, at least in part, to have driven the recent and rapid radiation of this species group (Henry et al., 2013). Crossing studies have found that mating song phenotypes in the *carnea*-group are genetically controlled by few loci of large effect (Henry et al. 2002), but to date, no loci underlying mating song phenotypes have been identified.

In this species complex multiple pairs of sympatric species vary primarily in their courtship signals. In two instances, pairs of species in this group have been recognized as distinct based upon mating signals: *Chrysoperla plorabunda* and *Chrysoperla adamsi,* which co-occur in western North America, and *Chrysoperla carnea* and *Chrysoperla pallida,* which co-occur in western Europe. All lack strong morphological and ecological differences (Henry et al., 1993; Henry et al., 2002). Individuals of each pair have been collected from the same fields at the same time and perform courtship songs that are similar in structure except for a consistent, major difference in one critical song feature, volley period. The mating signal of species in the *carnea*-group may be composed of one or more bursts of sound. Volley period is the length of a single

burst of sound **(Fig 1)**. Volley period difference is critical for reproductive isolation between the species in each of these two pairs (Henry et al., 1993; Henry et al., 2002).

As both of these species pairs can be coerced into producing fertile hybrid offspring in the laboratory, we have the potential to associate the volley period with genomic regions using QTL mapping. Together with divergence scans, this approach can identify the genomic basis of mating song and point to those genomic regions underlying repeated and rapid speciation seen across the *Chrysoperla carnea* species-group.

Herein, we identified the genomic underpinnings of volley period, a trait important to reproductive isolation, and began to reveal the genomic architecture of speciation in the *Chrysoperla carnea*-group. We estimated the number of loci underlying volley period and tested for genetic linkage between volley period phenotype and preference. Using QTL mapping, we identified a region of the genome associated with volley period. Finally, we scanned for signals of genomic divergence surrounding those loci as a potential second line of evidence to support our QTL mapping.

Methods

Collecting and laboratory rearing

Chrysoperla plorabunda (Fitch) and *Chrysoperla adamsi* (Henry et al.) cross parents were collected in western Oregon, USA in 2015. The wild-caught *Chrysoperla adamsi* female was collected from shrubby vegetation near the top of Mary's Peak, OR. The wild-caught *Chrysoperla plorabunda* male was collected from a clover field along a road in Benton, OR. *Chrysoperla carnea* (Stephens) and *Chrysoperla pallida* (Henry et al.) were collected from the same rooftop light in Zürich, Switzerland in 2016.

All lacewings were laboratory-reared on a long day light cycle (16 hours light, 8 hours

dark) at 25 ± 1°C. Adults were kept in clear plastic 8 oz cups with a hole in the bottom for a cotton wick. Rearing cups were stacked inside a second clear plastic cup containing a water reserve, and covered by a petri dish lid. A 2:1:1 mixture of Wheast, honey, and water was provided *ad libitum* for adults. Larvae were reared in clear plastic 1 oz cups with snap-on lids. Sterile *Ephestia kuchinella* eggs were provided to the larvae *ad libitum*.

Hybrid phenotypes

Mating signals for all individuals were recorded in the laboratory at 25 ± 1°C using an optical microphone (see Henry et al., 2013 Figure 1). The wild-caught parents were identified to species by Charles Henry based upon mating signal. A single pair of wild-caught parents was induced to mate by confinement together in a single rearing container. F_1 hybrid offspring were reared to adulthood and brother-sister mated to produce F_2 hybrids. The most productive *C. adamsi* x *C. plorabunda* cross produced 78 F_2 offspring. The most productive *C. pallida* x *C. carnea* cross produced 51 F_2 offspring.

Inheritance of song phenotype and song preference

The minimum number of genetic components underlying volley period phenotype was estimated from the variance in the phenotype data from all crosses generated for each species pair (Lande, 1981). Mating signal preference was assessed in F_2 offspring from *C. pallida* x *C. carnea* crosses by measuring the fidelity of response to computer-generated mating signals characterized by volley periods spanning the range of period lengths between the two parental species. Playbacks were conducted through an electronic shaker plate. The ten computer-generated songs with different volley periods were played five times each, for a total of fifty signals presented to each F_2 offspring individual in a random order. The weighted average of all volley periods to which the individual responded was used as a proxy for preferred period.

Individuals that responded to two or fewer of the fifty total signals played back were excluded from the analysis. The relationship between the signal an individual F_2 produced and the signal it preferred was assessed with Pearson's correlation coefficient.

Sequencing and RAD locus construction

Sequencing was completed only for the *C. adamsi* x *C. plorabunda* cross. DNA was extracted using a Qiagen DNeasy blood and tissue kit from the two parents founding the line, 13 F_1 individuals, and 78 F_2 individuals. A reduced representation sequencing (RAD sequencing) library of all individuals was prepared by Floragenex with the PstI cutting enzyme and sequenced on two lanes of an Illumina HiSeq at the University of Oregon. Single-end 150 base pair reads were demultiplexed with the process_radtags scripts in the STACKS pipeline. Quality control and adapter content removal was performed with TRIMMOMATIC (Bolger et al., 2014). Reads were mapped against the reference assembly of *Chrysoperla carnea* (SUB6529655) with BWA (Li & Durbin, 2009), and RAD loci were constructed with default parameters in STACKS version 1.4 (Catchen et al., 2013). Two F_2 individuals were excluded from the analysis, one based upon deviation from expected lacewing GC content distribution in the reads (mean > 50%), and the second for high levels of missing data after the removal of significant adapter contamination. Loci with significant deviations ($p < 0.05$) from a X^2 normality test (1:2:1) or with excessive missing data, along with 52 identical loci, were pruned using JoinMap 5 (Van Oojien, 2018).

Linkage map construction and QTL mapping

A linkage map was constructed using the regression mapping algorithm (Stam, 1993) in JoinMap 5 (Van Oojien, 2018). Associations between linkage groups and song phenotypes were detected with interval mapping (Lander & Botstein, 1989) as implemented in MapQTL 6 (Van

Ooijen, 2009). The significance threshold of 4.2 was determined by a permutation test in MapQTL 6 (Van Ooijen, 2009).

Association of linkage groups and genomic scaffolds for divergence estimates and candidate gene identification

Genomic scaffolds were associated with the linkage map using default settings in Chromonomer (Catchen et al., 2020). RAD sequencing data from six wild-caught *C. adamsi* and five *C. plorabunda* (phylogeny chapter data) were aligned to the scaffolds in the third linkage group with BWA (Li & Durbin, 2009), and RAD loci were constructed with STACKS 2.41 (Rochette et al., 2019). Population divergence, measured as $F_{ST,}$ was calculated for all variable sites where sequence was present in at least two individuals for each species, and smoothed F_{ST} was calculated over windows of 150 kilobases. The significance of F_{ST} values was determined with Fisher's exact test. Genes in the region strongly associated with volley period were identified and their annotations were parsed.

Results

Phenotypes

Volley periods for F_1 and F_2 offspring in both species pairs were intermediate to the parental phenotypes, showing greater phenotypic range in the F_2 phenotypes (**Fig 1**). The minimum number of genetic factors estimated to underlie volley period phenotype in the *C. adamsi* x *C. plorabunda* cross was 2.5. For the *C. carnea* x *C. pallida* cross the estimate was 5.7. Volley period phenotype and volley period preference of F_2 hybrids of *C. carnea* and *C. pallida* were strongly correlated, r = 0.71 (**Fig 2**).

Linkage map construction, QTL mapping and divergence scan

After filtering, 2,188 loci remained for two parents and 78 F_2 offspring of the *C. adamsi* x *C. plorabunda* cross. Six linkage groups with LOD scores of ten and that included more than ten markers were identified as likely corresponding to the six chromosomes known for *Chrysoperla carnea* (Naville, & Jacques de Beaumont, 1936). A large portion of linkage group three is strongly associated with volley period phenotype (**Fig 3**). In several regions of linkage group three the LOD score is above the significance threshold determined by the permutation test.

1,862 of the 40,501 genomic scaffolds representing ~12% of total length were able to be associated with the linkage map due to marker density and fragmentation of the genome. In the scaffolds associated with the linkage map, there were loci with higher and lower divergence interspersed across the linkage groups. However, no clear signal of increased F_{ST} was seen on linkage group three surrounding the putative volley period-associated region **(Fig 4)**.

Candidate genes

The scaffolds linked to the putative volley period QTL contained a total of 348 predicted genes. Three of those genes were associated with the gene ontology term for courtship behavior (GO:0007619). One of those predicted genes was *doublesex*, a gene necessary for vibrational mating signal production in *Drosophila* (Dauwalder, 2011; Rideout et al., 2010; Yamamoto & Koganezawa, 2013). The two other predicted genes associated with courtship behavior in the volley period QTL were *Prospero* and a gene highly similar to *cap-specific mRNA (nucleoside-2'-O-)-methyltransferase 2*.

Discussion

We explored the underlying genomic source of reproductive isolation and speciation in a rapidly evolving lacewing group. Using QTL mapping for an F_2 cross between *Chrysoperla plorabunda* and *Chrysoperla adamsi* we identified a large genomic region associated with

mating song phenotype. In that region we find the gene *doublesex*. This sex determination gene is necessary for typical song production in *Drosophila* (Dauwalder, 2011; Rideout et al., 2010; Yamamoto & Koganezawa, 2013). *doublesex* expression during development is required for connection between thoracic interneurons and wing motor neurons responsible for song production (Shirangi et al., 2016). We also found other candidate genes associated with courtship behavior. Fine-scale QTL mapping or functional validation with RNAi or CRISPR/Cas9 will be necessary to confirm the relationship between our identified candidate genes and volley period.

Phenotype distributions in our crossing experiments suggest that we should expect more than 2.5 distinct regions to underlie volley period phenotype for this hybrid cross. Our QTL mapping identified a single large volley period-associated region. These findings align with previous estimates (Henry et al., 2002). Together with the prior literature our findings suggest that the genetic control of *carnea*-group mating signals may be simpler than the genetic control of mating signals in other insects (Blankers et al., 2019; Gleason, 2005; Greenspan & Ferveur, 2000).

Though we identified only one QTL region, other loci of small effect related to volley period are likely spread throughout the genome and undetected in this analysis. Our power to detect loci is dependent upon the effect size, and thus we expect a significant bias towards identifying only loci of large effect with this type of approach (Broman, 2001). Possibly, the large region and multiple peaks in our trait association mapping, and the multiple genes related to courtship behavior in that region, could indicate that multiple genes controlling volley period are located together in our QTL. However, the associated region may simply be large because fine-scale mapping is difficult with small numbers of recombination events and low marker density (Broman, 2001; Mackay et al., 2009).

Using divergence scans on a very small sample of wild-caught lacewings, we find some variants with elevated divergence in the volley period-associated genomic region. However, we do not find the large region of elevated F_{ST} we might expect to surround this trait critical to reproductive isolation. The divergence patterns observed here cannot further validate the putative region identified with QTL mapping.

We found that song phenotype and song preference are tightly linked in F_2 hybrids of *C. carnea* and *C. pallida*. This relationship post-recombination suggests that the two traits are controlled by regions of the genome that are very tightly linked, or alternatively controlled by the same locus or loci through pleiotropy. These findings do not agree with those of Noh & Henry (2015), an analysis that was performed on a smaller sample size and with a different mating cross within the same species group. Linkage between these traits could facilitate rapid speciation in these organisms by limiting recombination. Potentially, a single mutation event could simultaneously change both signal and preference, resulting in the immediate creation of a new reproductively isolated lineage.

We have made significant steps towards characterizing the genetic architecture of speciation in the *Chrysoperla carnea*-group. Our findings suggest that a small number of loci underlie volley period, a mating song feature critical to reproductive isolation. We identified a single genomic region underlying volley period and multiple candidate genes. Additionally, we found evidence of genetic linkage between volley period phenotype and preference.

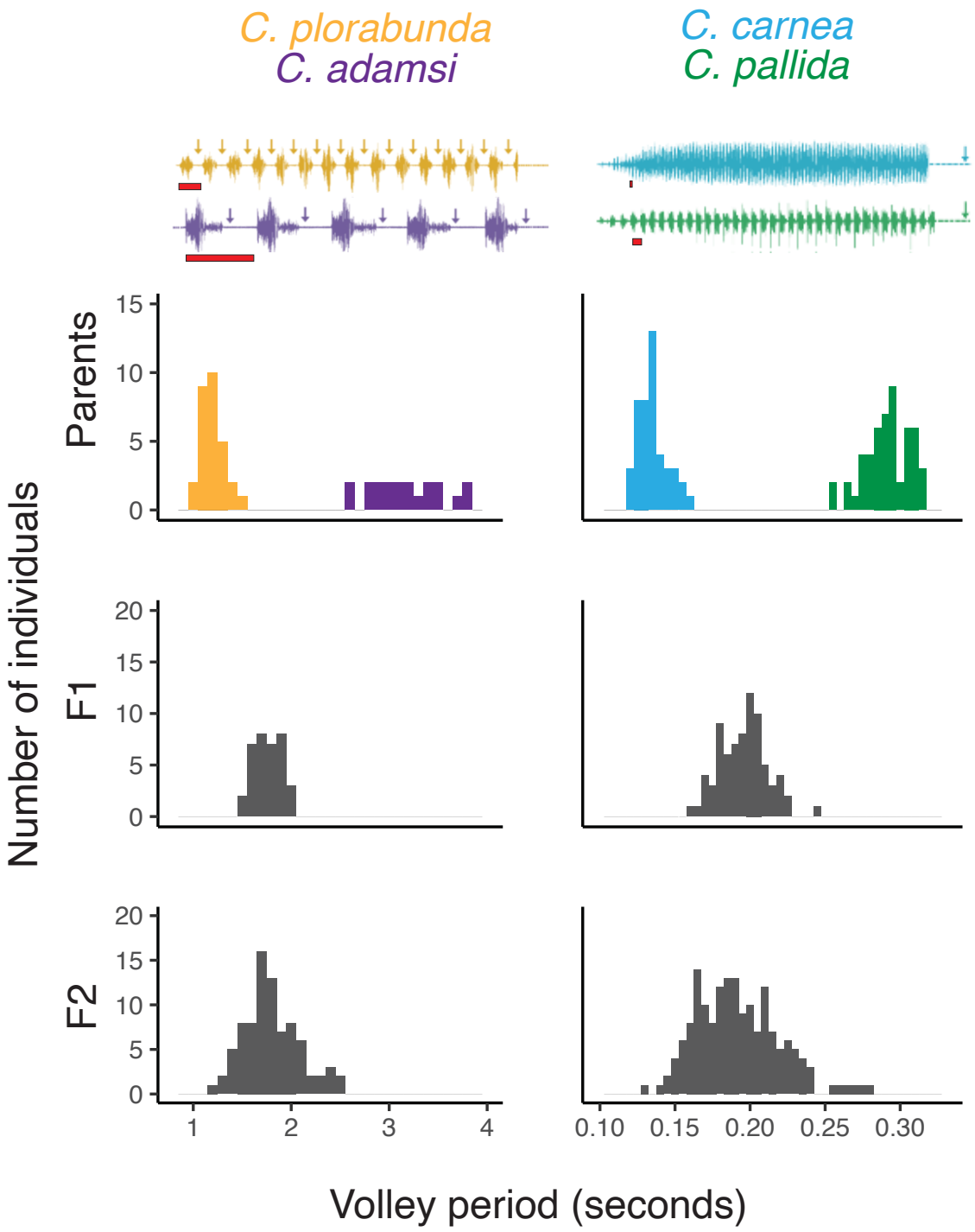

Fig 1. F_2 hybrid cross phenotypes. A. Four-second oscillograms of vibrational mating signal of the four parent species are displayed above volley period histograms for parental, F_1, and F_2 individuals from each cross. A single volley period for each of the parental species is indicated with a red line under the oscillogram. Arrows on the oscillogram indicate where a duetting partner would insert a response.

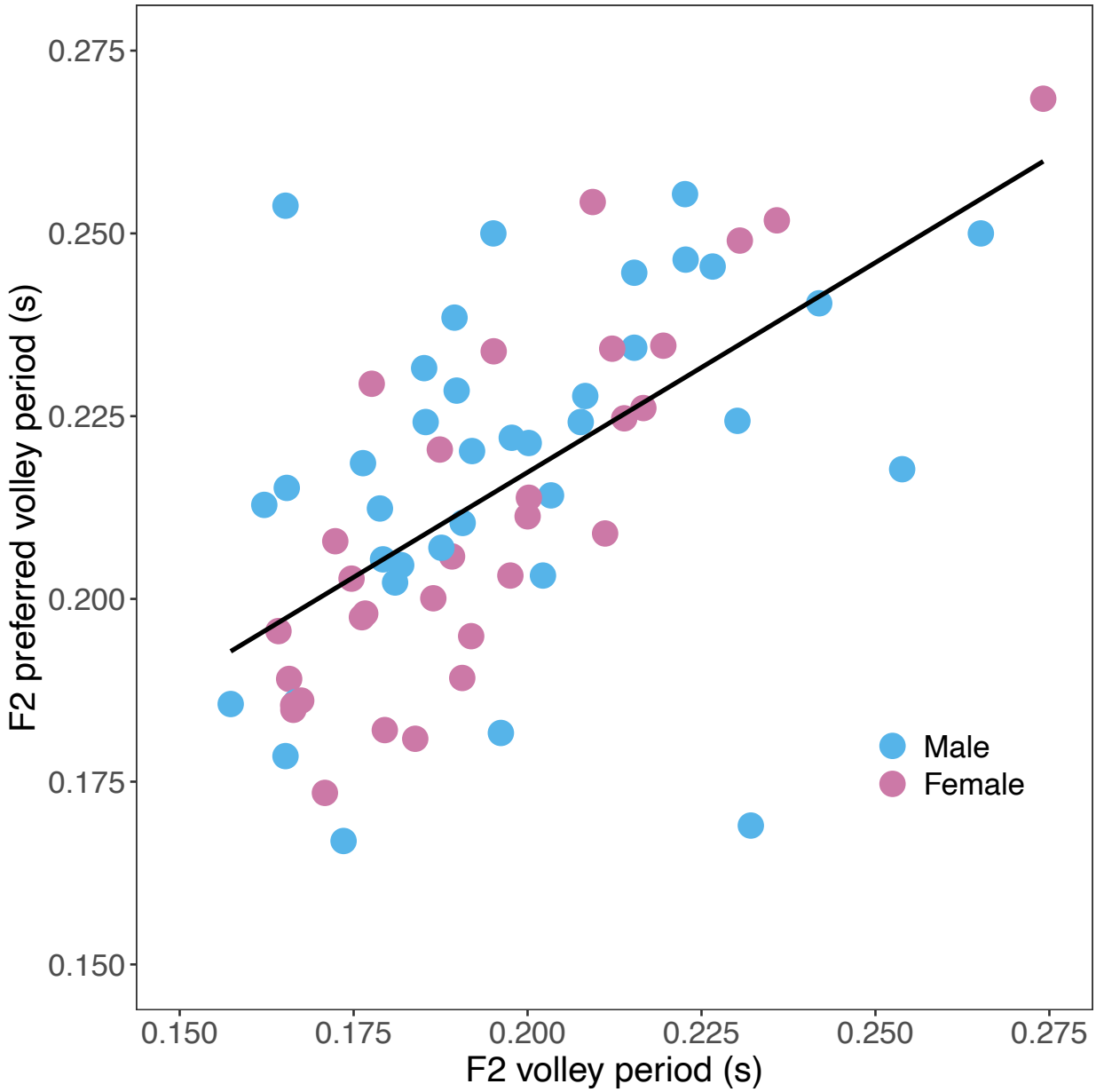

Fig 2. The correlation between the preferred volley period and the produced volley period of F_2 hybrid offspring of *C. carnea* and *C. pallida.*

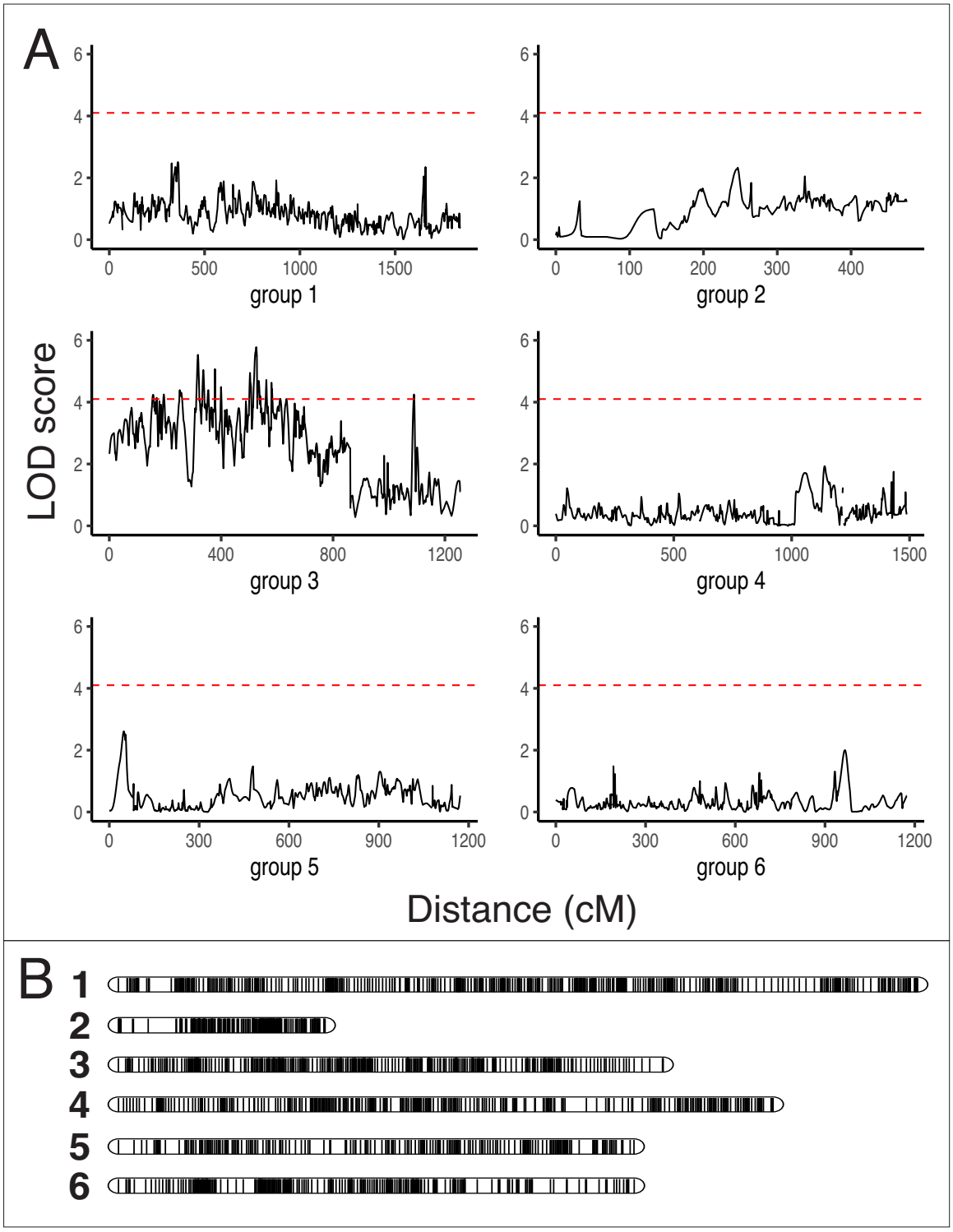

Fig 3. A. Linkage map constructed from a *Chrysoperla plorabunda* and *Chrysoperla adamsi* F_2 cross. B. Association of volley period with each of the six linkage groups as measured by LOD score. The significance threshold determined by permutation test is indicated by a red horizontal line.

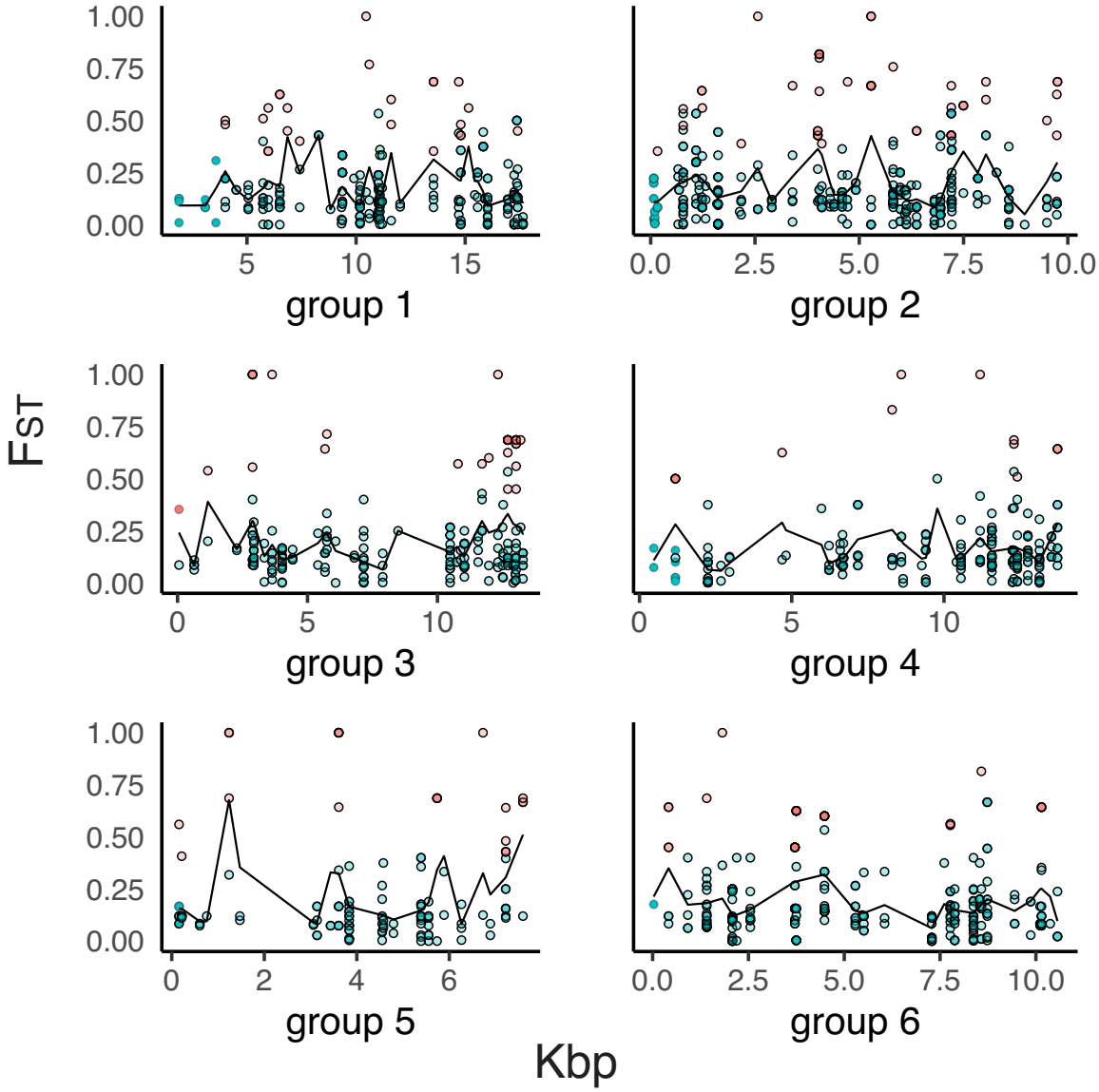

Fig 4. F_{ST} calculated for each variable site across the linkage map with point color indicative of significant test result. The plotted line is smoothed F_{ST}.

References

Blankers, T., Oh, K. P., & Shaw, K. L. (2019). Parallel genomic architecture underlies repeated sexual signal divergence in Hawaiian *Laupala* crickets. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1912), 20191479. https://doi.org/10.1098/rspb.2019.1479

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.

Broman, K. W. (2001). Review of statistical methods for QTL mapping in experimental crosses. *Lab Animal*, *30*(7), 44–52.

Campbell, C. R., Poelstra, J. W., & Yoder, A. D. (2018). What is Speciation Genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species. *Biological Journal of the Linnean Society*, *124*(4), 561–583. https://doi.org/10.1093/biolinnean/bly063

Catchen, J., Amores, A., & Bassham, S. (2020). Chromonomer: a tool set for repairing and enhancing assembled genomes through integration of genetic maps and conserved synteny. *BioRxiv*, 2020.02.04.934711. https://doi.org/10.1101/2020.02.04.934711

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. https://doi.org/10.1111/mec.12354

Coyne, J. A. (1992). Genetics and speciation. *Nature*, *355*(6360), 511–515. https://doi.org/10.1038/355511a0

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, *23*(13), 3133–3157. https://doi.org/10.1111/mec.12796

Dauwalder, B. (2011). The roles of fruitless and doublesex in the control of male courtship. In N. Atkinson (Ed.), *International Review of Neurobiology* (Vol. 99, pp. 87–105). Academic Press. https://doi.org/10.1016/B978-0-12-387003-2.00004-5

Gleason, J. M. (2005). Mutations and natural genetic variation in the courtship song of *Drosophila*. *Behavior Genetics*, *35*(3), 265–277. https://doi.org/10.1007/s10519-005-3219-y

Greenspan, R. J., & Ferveur, J.-F. (2000). Courtship in *Drosophila*. *Annual Review of Genetics*, *34*(1), 205–232. https://doi.org/10.1146/annurev.genet.34.1.205

Henry, C. S., Brooks, S. J., Duelli, P., & Johnson, J. B. (2002). Discovering the true *Chrysoperla carnea* (Insecta: Neuroptera: Chrysopidae) using song analysis, morphology, and ecology. *Annals of the Entomological Society of America*, *95*(2), 172–191. https://doi.org/10.1603/0013-8746(2002)095[0172:DTTCCI]2.0.CO;2

Henry, C. S., Brooks, S. J., Duelli, P., Johnson, J. B., Wells, M. M., & Mochizuki, A. (2013). Obligatory duetting behaviour in the *Chrysoperla carnea*-group of cryptic species (Neuroptera: Chrysopidae): its role in shaping evolutionary history. *Biological Reviews*, *88*(4), 787–808. https://doi.org/10.1111/brv.12027

Henry, C. S., Wells, M. L. M., & Holsinger, K. E. (2002). The inheritance of mating songs in two cryptic, sibling lacewing species (Neuroptera: Chrysopidae: *Chrysoperla*). In *Genetics of*

Mate Choice: From Sexual Selection to Sexual Isolation (Vol. 1–Book, Section, pp. 269–289). Springer.

Henry, C. S., Wells, M. M., & Pupedis, R. J. (1993). Hidden taxonomic diversity within <i>Chrysoperla plorabunda<\i> (Neuroptera: Chrysopidae): two new species based on courtship songs. *Annals of the Entomological Society of America*, *86*(1), 1–13. https://doi.org/10.1093/aesa/86.1.1

Kronforst, M. R., Young, L. G., Kapan, D. D., McNeely, C., O'Neill, R. J., & Gilbert, L. E. (2006). Linkage of butterfly mate preference and wing color preference cue at the genomic location of wingless. *Proceedings of the National Academy of Sciences*, *103*(17), 6575–6580. https://doi.org/10.1073/pnas.0509685103

Lande, R. (1981). The minimum number of genes contributing to quantitative variation between and within populations. *Genetics*, *99*(3–4), 541–553.

Lander, E. S., & Botstein, D. (1989). Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, *121*(1), 185–199.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Mackay, T. F. C., Stone, E. A., & Ayroles, J. F. (2009). The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics*, *10*(8), 565–577. https://doi.org/10.1038/nrg2612

Mullen, S. P., & Shaw, K. L. (2014). Insect speciation rules: Unifying concepts in speciation research. *Annual Review of Entomology*, *59*(1), 339–361. https://doi.org/10.1146/annurev-ento-120710-100621

Naville, André, & Jacques de Beaumont. (1936). Recherches sur les chromosomes des Névroptères. *Archives d'Anatomie Microscopique*, *32*, 271–302.

Noh, S., & Henry, C. (2015). Speciation is not necessarily easier in species with sexually monomorphic mating signals. *Journal of Evolutionary Biology*, *28*(11), 1925–1939.

Nosil, P., & Feder, J. L. (2012). Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1587), 332–342. https://doi.org/10.1098/rstb.2011.0263

Nosil, P., & Schluter, D. (2011). The genes underlying the process of speciation. *Trends in Ecology & Evolution*, *26*(4), 160–167. https://doi.org/10.1016/j.tree.2011.01.001

Rideout, E. J., Dornan, A. J., Neville, M. C., Eadie, S., & Goodwin, S. F. (2010). Control of sexual differentiation and behavior by the doublesex gene in Drosophila melanogaster. *Nature Neuroscience*, *13*(4), 458–466. https://doi.org/10.1038/nn.2515

Rochette, N. C., Rivera-Colón, A. G., & Catchen, J. M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, *28*(21), 4737–4754. https://doi.org/10.1111/mec.15253

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G.-P., Bank, C., Brännström, Å., Brelsford, A., Clarkson, C. S., Eroukhmanoff, F., Feder, J. L., Fischer, M. C., Foote, A. D., Franchini, P., Jiggins, C. D., Jones, F. C., …

Widmer, A. (2014a). Genomics and the origin of species. *Nature Reviews Genetics*, *15*(3), 176–192. https://doi.org/10.1038/nrg3644

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G.-P., Bank, C., Brännström, Å., Brelsford, A., Clarkson, C. S., Eroukhmanoff, F., Feder, J. L., Fischer, M. C., Foote, A. D., Franchini, P., Jiggins, C. D., Jones, F. C., … Widmer, A. (2014b). Genomics and the origin of species. *Nature Reviews Genetics*, *15*(3), 176–192. https://doi.org/10.1038/nrg3644

Shaw, K. L., & Lesnick, S. C. (2009). Genomic linkage of male song and female acoustic preference QTL underlying a rapid species radiation. *Proceedings of the National Academy of Sciences*, *106*(24), 9737–9742. https://doi.org/10.1073/pnas.0900229106

Shirangi, T. R., Wong, A. M., Truman, J. W., & Stern, D. L. (2016). Doublesex regulates the connectivity of a neural circuit controlling *Drosophila* male courtship song. *Developmental Cell*, *37*(6), 533–544. https://doi.org/10.1016/j.devcel.2016.05.012

Stam, P. (1993). Construction of integrated genetic linkage maps by means of a new computer package: Join Map. *The Plant Journal*, *3*(5), 739–744. https://doi.org/10.1111/j.1365-313X.1993.00739.x

Templeton, A. R. (1981). Mechanisms of speciation - A populations genetic approach. *Annual Review of Ecology and Systematics*, *12*(1), 23–48. https://doi.org/10.1146/annurev.es.12.110181.000323

Van Ooijen, J. W. (2009). *MapQTL 6, Software for the mapping of quantitative trait loci in experimental populations of diploid species* (Version 6) [Computer software]. Kyazma B. V.

Van Oojien, J. W. (2018). *Joinmap 5, software for the calculation of genetic linkage maps in experimental populations of diploid species* (Version 5) [Computer software]. Kyazma B. V. https://www.scienceopen.com/document?vid=b159fadd-708e-4c3e-9c91-bd395f88b847

Wiley, C., Ellison, C. K., & Shaw, K. L. (2012). Widespread genetic linkage of mating signals and preferences in the Hawaiian cricket *Laupala*. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1731), 1203–1209. https://doi.org/10.1098/rspb.2011.1740

Wilkins, M. R., Seddon, N., & Safran, R. J. (2013). Evolutionary divergence in acoustic signals: causes and consequences. *Trends in Ecology & Evolution*, *28*(3), 156–166. https://doi.org/10.1016/j.tree.2012.10.002

Wu, C. I., & Ting, C. T. (2004). Genes and speciation. *Nature Reviews Genetics*, *5*(2), 114–122. https://doi.org/10.1038/nrg1269

Xu, M., & Shaw, K. L. (2019). Genetic coupling of signal and preference facilitates sexual isolation during rapid speciation. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1913), 20191607. https://doi.org/10.1098/rspb.2019.1607

Yamamoto, D., & Koganezawa, M. (2013). Genes and circuits of courtship behaviour in *Drosophila* males. *Nature Reviews Neuroscience*, *14*(10), 681–692. https://doi.org/10.1038/nrn3567